AMERICAN
PSYCHOLOGICAL
ASSOCIATION

Volume 110 February 2018
Number 2

Published eight times ISSN 0022-0663

Journal of Educational Psychology

www.apa.org/pubs/journals/edu

The main purpose of the *Journal of Educational Psychology*® is to publish original, primary psychological research pertaining to education across all ages and educational levels. A secondary purpose of the *Journal* is the occasional publication of exceptionally important theoretical and review articles that are pertinent to educational psychology.

Single Issues, Back Issues, and Back Volumes: For information regarding single issues, back issues, or back volumes, write to Order Department, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242; call 202-336-5600 or 800-374-2721; or visit www.apa.org/pubs/journals/subscriptions.aspx

Manuscripts: Submit manuscripts electronically through the Manuscript Submissions Portal found at www.apa.org/pubs/journals/edu according to the Instructions to Authors found elsewhere in this issue (see table of contents). Correspondence regarding manuscripts should be sent to the Editor, Steve Graham, at steve.graham@asu.edu.

Electronic Access: APA members who subscribe to this journal have automatic access to all issues of the journal in the PsycARTICLES® full-text database. See http://my.apa.org/access.html.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

APA Journal Staff: Rosemarie Sokol-Chang, PhD, *Publisher, APA Journals*; Mare Meadows, *Managing Director*; Richard Pepple, *Journal Production Editor*; Cheryl Johnson, *Peer Review Coordinator*; Jodi Ashcraft, *Director, Advertising Sales and Exhibits.*

Journal of Educational Psychology® (ISSN 0022-0663) is published eight times (January, February, April, May, July, August, October, November) in one volume per year by the American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242. Subscriptions are available on a calendar year basis only (January through December). The 2018 rates follow: *Nonmember Individual:* $263 Domestic, $305 Foreign, $327 Air Mail. *Institutional:* $1,020 Domestic, $1,097 Foreign, $1,121 Air Mail. *APA Member:* $127. *APA Student Affiliate:* $79. Write to Subscriptions Department, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242. Printed in the U.S.A. Periodicals postage paid at Washington, DC, and at additional mailing offices. POSTMASTER: Send address changes to *Journal of Educational Psychology,* American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.

Effective with the 1986 volume, this journal is printed on acid-free paper.

Journal of

Educational Psychology®

www.apa.org/pubs/journals/edu

February 2018

Volume 110
Number 2

In Celebration of Educational Psychology

147 Past as Prologue: Educational Psychology's Legacy and Progeny
Patricia A. Alexander

163 Educational Psychology: A Future Retrospective
Karen R. Harris

174 Educational Psychology's Past and Future Contributions to the Science of Learning, Science of Instruction, and Science of Assessment
Richard E. Mayer

Reading

180 Does Writing System Influence the Associations Between Phonological Awareness, Morphological Awareness, and Reading? A Meta-Analysis
Yufang Ruan, George K. Georgiou, Shuang Song, Yixun Li, and Hua Shu

Learning

203 Testing Prepares Students to Learn Better: The Forward Effect of Testing in Category Learning
Hee Seung Lee and Dahwi Ahn

218 Anthropomorphism in Decorative Pictures: Benefit or Harm for Learning?
Sascha Schneider, Steve Nebel, Maik Beege, and Günter Daniel Rey

233 How Affective Charge and Text–Picture Connectedness Moderate the Impact of Decorative Pictures on Multimedia Learning
Sascha Schneider, Jonathan Dyrna, Luis Meier, Maik Beege, and Günter Daniel Rey

250 Animated Pedagogical Agents as Aids in Multimedia Learning: Effects on Eye-Fixations During Learning and Learning Outcomes
Fuxing Wang, Wenjing Li, Richard E. Mayer, and Huashan Liu

Motivation

269 Daily Autonomy Supporting or Thwarting and Students' Motivation and Engagement in the High School Science Classroom
Erika A. Patall, Rebecca R. Steingut, Ariana C. Vasquez, Scott S. Trimble, Keenan A. Pituch, and Jen L. Freeman

289 The Moderating Role of Popular Peers' Achievement Goals in 5th- and 6th-Graders' Achievement-Related Friendships: A Social Network Analysis
Lydia Laninga-Wijnen, Allison M. Ryan, Zeena Harakeh, Huiyoung Shin, and Wilma A. M. Vollebergh

(*Contents continue*)

Other

308 Call for Papers - A Focused Collection of Qualitative Studies in the Psychological Sciences: Reasoning and Participation in Formal and Informal Learning Environments
179 E-Mail Notification of Your Latest Issue Online!
iii Instructions to Authors
ii Subscription Order Information

Journal of Educational Psychology
2018, Vol. 110, No. 2, 147–162

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000200

Past as Prologue: Educational Psychology's Legacy and Progeny

Patricia A. Alexander
University of Maryland

On the occasion of the 125th anniversary of the American Psychological Association, the legacies and progenies of the discipline of educational psychology are explored. To capture those legacies, transformational and influential contributions by educational psychologists to schools and society are described as key themes. Those themes entail: the "psychologizing" of education, engagement in interdisciplinary and cross-disciplinary inquiry, a focus on learning as a core construct, an investment in measurement and an appreciation of human variability, and a search for evidence-based approaches and practices that work. To project forward, those same thematic areas are revisited 25 years from now as the means of speculating on educational psychology's future contributions to schools and society. In both the case of the legacies and progenies, potential difficulties or particular challenges are also considered.

Educational Impact and Implications Statement

For over 125 years, members of the educational psychology community have made untold contributions to society. In this article, those untold contributions are distilled into five areas of influence that serve as the discipline's enduring legacy: the "psychologizing" of education; interdisciplinary and cross-disciplinary inquiry; learning; individual differences and their measurement; and evidence-based approaches and practices. The article also offers a glimpse into what may well be educational psychology's future contributions, compelled by dramatic changes already on the horizon.

Keywords: assessment, cognition, individual differences, technology

In 2017, the American Psychological Association and the field of educational psychology celebrate their 125th birthday—true milestones. These events are milestones not only in the metaphorical sense but in significant and concrete ways. Dating back to ancient Rome and the Appian Way, milestones served a very functional purpose. Their function was to allow travelers to mark how far they had progressed in their journey. Certain markers, such as the Golden Milestone in ancient Rome, were of particular significance because they were the zero points from which all directions were to be measured—a point of reckoning. I would like to use the occasion of this quasquicentennial to create such a point of reckoning for educational psychology—to mark not only where the field has come but also where it is heading. In effect, this milestone can signify how much ground the community and its members have traversed in the past 125 years, especially in terms of its many contributions to schools and society. For another, this milestone can signal new directions to be pursued and the new terrains to explore.

In this treatise, which is part retrospective and part prospective, I will undertake the role of historian documenting what I regard as several of educational psychology's most notable contributions to schools and society. Although I label these as "contributions," I will likewise consider the complications that have arisen as a consequence of these developments. Next, I will assume the role of prognosticator. In this capacity, I will project forward in time, envisioning paths and destinations that members of this community may pursue in the years to come. Of course, I operate under no illusion that either role of historian or prognosticator will be performed flawlessly. Yet, I trust that the legacies and progenies I discuss will fuel debate and discussion, for it would be of limited value to engage in this analysis unless reflection and critique ensue.

The Legacies

It was no simple task to distill the incredible number of contributions that the members of the educational psychology community have made to schools and society down to only a few. But it was a task that I judged to be of particular value, since it is rare for a field to have the opportunity to bear witness to its own accomplishments or to recognize the influence it has exerted on educational policies and practices. The contributions to which I refer are, by necessity, somewhat thematic in character, allowing for a broader consideration of the legacy they represent, rather than individual innovations or insights. Further, what I will raise as the significant and enduring legacies of educational psychology over the past 125 years may not, on the surface, appear earthshaking to those who have been members of the community for years. Perhaps that is because those embedded in the discipline and actively engaged in research may not have the time or the inclination to delve into the distant past or to weigh the academic, social, and

This article was published Online First March 23, 2017.

Correspondence concerning this article should be addressed to Patricia A. Alexander, Department of Human Development and Quantitative Methodology, College of Education, University of Maryland, College Park, MD 20742-1131. E-mail: palexand@umd.edu

political implications of what has or is transpiring within the discipline.

With these cautionary notes in mind, I contend that the following themes represent transformational and influential contributions that the educational psychology community has made over the course of its history:

- the psychologizing of education,
- interdisciplinary and cross-disciplinary inquiry,
- learning as a theoretical and empirical core,
- investment in measurement and an appreciation of human variability, and
- search for effective evidence-based approaches and practices.

The Psychologizing of Education

Psyche, from Latin, represents an animated spirit, or the soul, mind, or invisible force that directs human thought and action. When the discipline of educational psychology was conceived, it was with the intent to bring that animated spirit into the practice of education, and to do so through scientific inquiry (Bredo, 2003). While it may be hard for us to grasp contemporarily, it was regarded as unconventional and controversial at the turn of the 20th century that those who were devoted to critical and systematic inquiry—whether through a more philosophical or empirical lens—would turn their attention to education as a legitimate venue for investigation (Berliner, 2006; Pajares, 2003). Yet, that is precisely what occurred, and it was largely the pragmatists—Dewey, James, and Peirce—who were catalysts within the United States. The combined weight of John Dewey's passionate discourse, William James's incredible knowledge base, and Charles Peirce's mathematical and logical mind was an undeniable force that opened the door to the discipline we now call educational psychology.

For example, in *The Child and the Curriculum*, Dewey (1902) writes passionately of the need to *psychologize* school content. Describing the intentions of a teacher, Dewey (1902) writes:

> His problem is that of inducing a vital and personal experiencing. Hence, what concerns him, as teacher, is the ways in which that subject may become a part of experience; what there is in the child's present that is usable with reference to it; how such elements are to be used; how his own knowledge of the subject-matter may assist in interpreting the child's needs and doings, and determine the medium in which the child should be placed in order that his growth may be properly directed. He is concerned, not with the subject-matter as such, but with the subject-matter as a related factor in a total and growing experience. Thus to see it is to psychologize it. (p. 23)

Dewey's push to look deep inside classrooms, teacher practices, and subject matter and to extract meaningful principles that could reshape the educational experience remains characteristic of today's educational psychologists.

Similarly, in *Talks to Teachers on Psychology*, a compilation of lectures given at Harvard to Cambridge teachers, William James set out to make the emerging science of psychology accessible to practicing educators. In those lectures, James (1899) wrote about the integration of the art of teaching with the science of psychology. However, he cautioned that there was no direct path between the art of teaching and psychology, which he defined as the "science of the mind's laws" (James, 1899, p. 7). Rather, what James (1899) perceived as essential was "an intermediary inventive mind" that "must make the application, by using its originality" (p. 8). It was for educational psychologists to serve as those "intermediary inventive minds."

Although the influence of Peirce may not parallel that of his more popular colleagues, Dewey and James, we should not underestimate his contributions to educational psychology. What Peirce brought to the thoughts and writings of Dewey and James was an abiding concern for clarity, logic, reasoning, and "a fixation of belief." As Peirce (1877) argued: "The irritation of doubt causes a struggle to attain a state of belief. I shall term this struggle *inquiry*, though it must be admitted that this is sometimes not a very apt designation" (p. 5). We still see Peirce's influence in the literatures on epistemic beliefs, conceptual change, persuasion, and in the emergence of situated models of learning that rely heavily on perception.

We also hear echoes of Dewey's, James's, and Peirce's call to wed the psyche with the real concerns of education in the first editorial of the *Journal of Educational Psychology* published by the editors Bagley, Bell, Seashore, and Whipple (1910):

> The editors of this Journal believe that there is equal need of a "middle-magazine"—of a journal that shall afford a common meeting ground for the psychologist and the educator. We seek to supply the worker in the laboratory with a channel for the promulgation of those results of his investigation of mental life that bear, directly or indirectly, upon the problems of teaching, and we seek to enlist and stimulate the interest of schoolmen in the discussion of the varied and highly important problems of education that have psychological bearing. We regard, then, this Journal as a clearinghouse for the exchange of information upon all that concerns the relation of psychology to education. (p. 1)

Thus, from the time of its inception, educational psychology and those who call themselves members of this community have accepted the mantle of "intermediary inventive minds" intent on bringing scientific evidence into the realm of education.

As I stated at the outset, I do not want to simply give voice to these legacies in a nonjudgmental manner, I want to briefly consider any residual effects from these happenings. Although there is no question that the investment in education as a venue for scientific investigation has resulted in untold advancements in the past 125 years, there is one undeniable consequence—the waxing and waning relationship between educational psychology researchers and educational practitioners (Alexander, Murphy, & Greene, 2012). No one captured this concern for the relevance of educational psychology to educational policy and practice better than David Berliner in his opening chapter to the 2006 *Handbook of Educational Psychology*:

> To see ourselves, instead, as psychologizing about the problems and issues of education is different in subtle but important ways from simply being a middle-man. It is the difference between having a hammer and seeing the work in terms of nails that we might put in, versus understanding the goals of the architect, the function the structure is to serve, and the behavior of the people who will inhabit the structure. It is the difference between bringing behavioral psychology or self-efficacy theory or mastery learning to teachers having trouble getting high levels of achievements from some students, versus trying to understand what it is about this mix of teacher, student, curriculum, and setting that might be better understood through a strong grounding in psychology. (Berliner, 2006, p. 23)

To many practitioners then, as now, there can be the sense that educational psychologists treat them as uninformed novices who operate from flawed common sense rather than reasoned discernment (Broekkamp & van Hout-Wolters, 2007). Conversely, there are those within the educational psychology community who feel that their understandings and evidentiary-based recommendations fall upon deaf ears or that educational policies overlook critical empirical evidence (Berliner, 2008).

Perhaps the relationship between educational psychology researchers and school practitioners and policymakers will never be entirely harmonious. However, it is an association that remains worthy of whatever efforts toward reconciliation are required. Minimally, it seems essential for those identifying as educational psychologists and seeking to forge meaningful relationships with members of educational institutions to understand the culture of those institutions and the pressing concerns that school leadership and practicing teachers face daily.

Interdisciplinary and Cross-Disciplinary Inquiry

Whether by design or by necessity, educational psychology, from its inception, was interdisciplinary—a gathering of scholars from such diverse disciplines as philosophy, psychology, medicine, and mathematics. What bound them together was the belief in the value of scientific inquiry in matters pertaining to education, teaching, and learning (Bagley et al., 1910). Moreover, the interdisciplinary character of educational psychology conveyed the belief that the complexity of problems pertaining to education demanded insights from multiple vantage points. The very title of the premier journal for the community, the *Journal of Educational Psychology*, included the identifiers *experimental pedagogy*, *child physiology and hygiene*, and *educational statistics* to reflect that interdisciplinarity.

One of the lessons that the distant past of educational psychology establishes—and that can inform future research endeavors—is that interdisciplinary and cross-disciplinary inquiry seemingly arises organically when the nature or complexity of the problems to be addressed demand it. As a case in point, the power of interdisciplinary inquiry is evident in the insights that have been forthcoming about learning and performance in academic domains. In effect, the strong foci on mathematics education, science education, social studies education, and the like within educational psychology have come from the marriage of expertise in the parent discipline (e.g., mathematics or history) with the expertise in learning, teaching, and assessment from education and psychology (de Jong, Linn, & Zacharia, 2013; Newton, Leonard, Evans, & Eastburn, 2012; Shulman, 1987; Wineburg, Martin, & Monte-Sano, 2012).

In addition, those invested in the study of beliefs or belief change represent a marriage of philosophy with psychology and education (Chinn, Buckland, & Samarapungavan, 2011; Murphy & Mason, 2006). Indeed, there have even been extensive literatures that have taken general constructs from philosophy or psychology, such as epistemic beliefs or self-efficacy, and investigated them in terms of specific fields (Lee, Lee, & Bong, 2014; Mason, Boscolo, Tornatora, & Ronconi, 2013; VanSledright & Maggioni, 2016). It is also evident that what begins as new cross-disciplinary forays into education, teaching, and learning have become normalized as identified fields of study, such as mathematics or science education, or developmental and counseling psychology (Farleyet al., 2015).

The interdisciplinary nature of educational psychology remains a hallmark of the field, although the contributing fields understandably have shifted over time (Alexander et al., 2012). Take the growing fields of neuroscience and neuropsychology as cases in point. Neuroscience and neuropsychology, which are making inroads into educational research, are interdisciplinary sciences that draw from such fields as cognitive and computer science, medicine, psychology, and mathematics. Contemporarily, the community of educational psychologists, or those who contribute to the relevant knowledge in educational psychology, routinely includes scholars trained as cognitive scientists, computer scientists, developmental psychologists, educational statisticians, neuroscientists, and the learning scientists. What qualifies these individuals as members of the educational psychology community writ large is not that they hold a degree from an established educational psychology program, but that they share in the mission of psychologizing educational experiences. They share in the goal of delving into the educational experience and translating it "into the immediate and individual experiencing within which it has its origin and significance" (Dewey, 1902, p. 21).

What possible side effects could result from such goal-oriented interdisciplinary studies of education, teaching, and learning? One particular side effect warrants discussion. That is, when individuals with diverse training and disciplinary roots come together to address educational issues and concerns, they bring their traditions with them, including potentially diverse lexicons, methodological practices, and standards of evidence. For instance, the literature is replete with examples of how the burgeoning of terminology has plagued educational research, in part because researchers have developed their own specialized lexicon to describe constructs and processes of interest (Dinsmore, Alexander, & Loughlin, 2008). The result is that similar terms have been applied to different phenomena or a multitude of terms and phrases have been created for the same phenomenon (Alexander, Schallert, & Hare, 1991; Murphy & Alexander, 2000). Such lexical diversity complicates the communication that must occur.

Yet, it is not solely the linguistic differences that come into play. There are also significant viewpoints to be reconciled. Those different viewpoints can translate into varied grain sizes in analysis and contrasting methodologies that must be navigated (Ercikan & Roth, 2009). For instance, those who collaboratively investigate text-processing problems might focus on neurological issues such as working memory or attentional control, whereas others might center on family and community influences. There are also those who devote their empirical energies in the classroom, orchestrating techniques or approaches intended to address those processing concerns. Yet, it is not only a matter of scope, but it is also the nature of the evidence sought in these investigations and the grain size of the resulting evidence that shape the data-analytic procedures that prove suitable (Tashakkori & Creswell, 2007; Winne & Baker, 2013).

There is no question that empirical evidence is a hallmark of educational psychology research currently, as it has been in years past (Alexander et al., 2012). Further, while there is an increased acceptance of varied research models, such as the embrace of *Pasteur's Quadrant* (Stokes, 2011), there still remains a compelling search for causality afforded by experimental investigations.

As members from other disciplines are welcomed into the educational psychology community, therefore, it would seem essential not only to find common ground with regard to theoretical and conceptual differences but also to negotiate and reconsider standards of design and evidence that have been longstanding.

Learning as a Theoretical and Empirical Core

Because concern for education has always encompassed a plethora of cognitive, behavioral, social, motivational, affective, and cultural dimensions manifested in varied contexts, it may seem challenging to identify any one construct as a particularly noteworthy contribution of educational psychology. However, it is my contention that one construct has long endured as a centerpiece of educational psychology theory and research—learning. I am joined in the contention by J. Carleton Bell (1914), the first managing editor of the *Journal of Educational Psychology*, who wrote:

> The central theme in all discussions of educational psychology will continue to be the learning process. This is of vital importance both to the scientist and to the educator, and, in spite of the work that has already been done, the ground in many directions has scarcely been broken. The JOURNAL can promise its readers many interesting contributions to our knowledge of the technique of learning. (p. 43)

Certainly, over the past 125 years, there have been ongoing debates as to the very nature of learning, as represented by the catalog of learning theories that have been forwarded, empirically tested, and evoked as the bases for instructional innovations and interventions. The contributions of these theories linger, even when their popularity or political currency fades. The mechanisms of learning they expose, the predictor and outcomes variables they emphasize, the way variables are operationalized, the terminology they add to the educational lexicon remain in the academic Ethernet.

Further, the formal educational systems operating in industrial and postindustrial societies are expected to produce positive and "relatively enduring change" for all those mandated to participate (Alexander, Schallert, & Reynolds, 2009, p. 186). This expectation has required the educational research community, in collaboration with curriculum experts, practitioners, and policymakers, to wrestle over the desired learning outcomes by age/grade and content domain. It has also called for the development of appropriate indicators of learning, as well as the construction and validation of measures of cognitive ability, achievement, social-emotional well-being, and more. This active pursuit of indicators and measures has also fueled the need for suitable measurement and statistical innovations that can convert resulting data into interpretable and impactful conclusions.

The concern for documented outcomes has also given rise to a rich literature in classroom-based techniques and interventions that facilitate learning (a) of valued knowledge or competencies, (b) for particular student populations, (c) for specific academic domains, and (d) within certain contexts (e.g., Cervetti, Kulikowich, & Bravo, 2015; McMaster et al., 2015; Star et al., 2015). Indeed, educational psychology has contributed much to understanding the multitude of factors within individuals, contexts, and tasks that influence what is learned and how learning transpires at any given moment in time and over time (Denissen, Zarrett, & Eccles, 2007; Marsh, 1990). A perusal of the three-volume *Handbook of Educational Psychology* edited by Harris, Graham, and Urdan (2012) is a tribute to those many contributions. For instance, there are those within the community who are particularly concerned with the effects that the overall classroom climate or peer relations play in students' academic success (Rubie-Davies, Flint, & McDonald, 2012; Wentzel, Barry, & Caldwell, 2004). Also, the rich literature in expertise development and the more recent interest in learning trajectories reflect an abiding concern for the path of learning over time (Alexander, 2003; Clements & Sarama, 2004; Nandagopal & Ericsson, 2012). Further, educational psychologists have long sought to solve the mystery of why some individuals appear to learn more effortlessly, quickly, more deeply, or more effectively than others and what cognitive, neurobiological, social, cultural, and affective/motivational forces seemingly underlie those differences (e.g., Hattie, 2008).

To that point, in seeking to encapsulate the research base undergirding the American Psychological Association's learner-centered psychological principles, Alexander, and Murphy (1998) concluded that "Learning, although ultimately a unique adventure for all, progresses through various common stages of development influenced by both inherited and experiential/environmental factors" (p. 36). These seemingly contrasting themes—what humans have in common educationally and what makes learning unique for each human—remain the tension with which educators and educational researchers must deal if they are to address learning for all students (Sternberg & Grigorenko, 2003). Related to this ongoing tension, educational psychologists have also been invested in ascertaining what techniques or approaches facilitate learning of particular content. For example, why does mathematical or scientific learning appear more challenging for some individuals or groups than for others (Oakes, 1990)? In light of their independent significance to the educational enterprise, however, I will elaborate further the independent contributions that individual and group differences, testing and measurement, and effective techniques and interventions have made—for better or worse—over the past 125 years. Before considering those notable contributions of educational psychology, however, it is necessary to explore potentially unwanted outcomes that arise from the focus on learning.

Let me offer one consequence that may not appear especially problematic on the surface but which nonetheless warrants the attention of those concerned with the positive and relative enduring changes we expect from schooling—the confounding of learning and achievement. I would assume that the need to raise this concern would not have been so pressing in generations past. However, the shifting landscape of schooling and the prevailing beliefs as to the purposes of learning and of teaching have brought this issue to the forefront.

More than a decade ago, Michelle Riconscente and I (Alexander & Riconscente (2005) wrote a provocative chapter for a volume dealing with the No Child Left Behind legislation that cast a shadow over American education. Our contention in that chapter was that achievement does not equate to learning and that the accountability push evident in policies and administrative practices of the time was undermining optimal learning in favor of test performance. In the intervening years, there has been little reason to assume that the conditions that sparked this treatise have improved dramatically. The emphasis on test performance seems to be as pervasive as ever and stands as a possible detriment to richer curricula or fuller considerations about learning, including stu-

dents' motivations for learning, or their perceptions as to what it means to be a good student (e.g., Berliner, 2011; Shepard, 2000).

Moreover, it is not solely the use of test scores to evaluate teachers (i.e., accountability) that is at issue. It is the even more pervasive effect of focusing students' attention on the scores or grades they receive in schools or on the narrow band of curriculum that will contribute directly to test performance (i.e., achievement; Berliner, 2011). I do not dispute some association between achievement and learning, but I clearly do not regard them as equivalent or even kindred concepts. For one, it is quite possible for students to score well or receive good grades without understanding the tested content deeply or critically. For another, there is ample evidence that certain students are not able to express what they know and can do in high-stakes testing conditions (Steele & Aronson, 1995). Thus, as both these circumstances suggest, achievement in the terms of test scores or grades is not a valid or reliable substitute for learning.

While I will deal with the aforementioned concern in more detail, let me conclude the current discussion with the observation that educational psychologists should minimally be sensitive to subtle but important distinctions between learning and achievement, and thoughtful in their choice of learning indicators employed in their investigations. They should also seek to communicate that distinction to policymakers, school administrators, and the general populace, so that test achievement is not treated as synonymous with the more central aim of education—learning.

Investment in Measurement and Embrace of Human Variability

Among the contributions that many educational researchers and practitioners would ascribe to educational psychology—for better or for worse—is the community's longstanding identification with measurement and assessment (Alexander et al., 2012). Over educational psychology's storied history that involvement has entailed the assessment of general cognitive constructs (e.g., intelligence, reasoning, creativity, or working memory), motivational/emotional factors (e.g., goal orientations, academic emotions, or self-efficacy), domain-specific aptitudes and achievements, and more (Pellegrino, 2004). Further, this enduring fascination runs the gamut, focusing on infancy through old age, targeting both typical and atypical populations, functioning within classroom as well as laboratory contexts, and encompassing researcher-developed local measures to high-stakes national assessments. As varied as these parameters are the purposes such measurements are intended to serve. For instance, measures have been crafted to unearth underlying general or specific capabilities; classify or categorize individuals into particular groups; diagnose specific processes or competencies; gauge achievement in academic domains; predict future academic or professional success; and expose underlying beliefs, predispositions, or attitudes. In essence, members of the educational psychology community are, by and large, quantifiers (Alexander et al., 2012).

Underlying these varied parameters and diverse purposes, however, there are certain circumstances that should be recognized. For one, since the time of Sir Francis Galton (1869) and his efforts to test his assumption that "eminence," (intelligence) was linked to genetics, there has been the expectation that variability is an intriguing and inevitable human condition. Moreover, so much of measurement theory and development, statistical analyses, and data interpretation rests on the notion proffered by Galton that there is a discernible pattern to human variability—the normal distribution. Where would educational psychology be as a discipline without the infamous bell curve? Galton's efforts to tie individual differences to human performance also resulted in the concept of correlation, which Pearson perfected years later (Walker, 1958), yet another basic tool in the educational researcher's toolkit.

But it is much more than an investment in the quantification of human variability that is a hallmark of the educational psychology community. There is the desire to truly understand the range of human variability, the forces that shape those differences, and the effects such variability may exert on learning and performance (Jonassen & Grabowski, 2012; Sternberg & Grigorenko, 2003). Perhaps more importantly, educational psychologists seek to determine what can be done educationally—whether in schools or the broader society—to recognize, support, and accommodate the differences within learning environments that will inevitably be encountered, and to intervene educationally when warranted (Odom et al., 2005). This fundamental desire to measure and address variability in all its manifestations drives much of the research in educational psychology now as it has in the past.

In the first volume of the *Journal of Educational Psychology* (1910), for example, there were articles devoted to the measurement of attention, mental fatigue, discrimination of brightness, pressure, sensitivity to pain, measures of retardation, vision span, physical growth, and general intelligence. Clearly, concerns for measurement were front and center in those early years of the discipline. Although the areas of particular interest within the community have obviously shifted and expanded over the past 125 years, the investment in testing and measurement and the concern for human variability is not less evident today. A survey of a recent volume of *Journal of Educational Psychology* (2016) included measures of achievement and performance for mathematics, science, history, English language learning, reading, and writing; teacher beliefs and motivations; topic interest and topic knowledge; social expectations and social climate; parental values; emotional exhaustion; self-concept; self-control; sexual orientation; reasoning; working memory, and more.

In that recent volume of *Journal of Educational Psychology*, there was also consideration of learners representing diverse populations (i.e., gender, culture, racial/ethnic diversity) and varying by age or level of achievement, as well as those who presented with specific learning or motivational profiles, or with particular social/emotional issues. What is also evident in the comparison of contemporary research to that of the distant past are the multiple measures typically administered in any one investigation, along with the sophisticated statistical procedures required to make sense of the resulting data. It is, therefore, an understatement to say that educational psychology research has become increasingly more complex in the questions asked and the measurement and statistics capabilities required to address those questions.

Yet, another noticeable and relevant difference that exists between past and contemporary research extends beyond the areas investigated, number of measures administered, and obvious sophistication of the statistical procedures employed. Specifically, in generations past, the focus of much of the research was centrally on the development of measures, be they cognitive, social, emo-

tional, motivational, or physiological in nature. By contrast, measures—while still critical to contemporary research—are the *means* by which weighty questions are tested and important policy and practice implications forwarded. This is more than a simple figure-ground issue. It speaks to the overall perspective that the present community holds about the place of test development and validation in the pursuit of larger educational goals.

Let me now turn to the potential side effects that merit examination when it comes to this theme of measurement and human variability. Specifically, within the last generation in particular, the focus on measurement and assessment within the educational system has expanded to the point that I contend that schools can justly be described as *test preparation institutions* rather than *bastions of learning* (Alexander, 2016a, in press). There has likewise been the concern that the expertise on assessment that resides within the educational psychology community has not always been well used in the policies and practices surrounding high-stakes or national tests (Darling-Hammond, 1992). I am not alone in my judgment that schools have invested so much of their time and energy on high-stakes testing and on accountability at the cost of optimal learning for all students (Berliner, 2011; Shepard, 2000).

I appreciate that there are movements underway to give states greater discretion with respect to indicators of their instructional performance as a way to reduce the accountability pressures (Every Student Succeeds Act, 2015). Nonetheless, there is a long way to go with regard to accountability for teachers and school leadership. Moreover, I see little comparable movement to reduce the presence of high-stakes assessments for the students who must bear the brunt of this testing obsession. This issue of the appropriate role of testing within the educational system will remain an ongoing concern for members of the educational psychology community who have a role to play in the development and validation of these high-stakes measures.

The second side effect pertains more to the members of the educational psychology community than the broader society. With the increasing complexity of contemporary research designs, the multitude of variables measured, and the growing sophistication needed to extract meaningful patterns from the resulting data, the substantive and methodological expertise required becomes daunting (Alexander et al., 2012). It is increasingly challenging to find individuals within the community who singularly possess both substantive and statistical/methodological knowledge and skills. This is even more the case with the rise in mixed methods studies that entail quantitative and qualitative methods intricately combined to address core research questions (DeCuir-Gunby, 2008). In some cases, these circumstances have fueled the wave of cross-disciplinary investigations populating journals, as well as the necessity of multiauthored publications. Perhaps this movement toward cross-disciplinary work between those with substantive and those with methodological expertise will prove a viable solution to the aforementioned challenges—but we must await the outcome.

Alexander et al. (2012) wrote expressly about the burgeoning of statistical and methodological techniques and the ramifications of that incremental development:

> To be sure, even a sampling of the literature illustrates this incremental advancement in methodologies. Classical techniques of handling missing data (i.e., listwise and pairwise deletion, mean imputation) are now shunned in favor of full information maximum likelihood and multiple imputation techniques. . . . Repeated measures and classroom environment effects that were analyzed in the past with ANOVAs (Analysis of Variance) are now modeled using multilevel analyses. . . . We see this trend as somewhat problematic in that as the number and sophistication of these techniques increase, educational psychologists are forced to spend time learning these new methodologies or risk becoming uncritical consumers of others' work. (p. 11)

Certainly the development of these advancements in statistical and methodological techniques have proven invaluable in addressing increasingly complex, multifaceted, and interdisciplinary questions of importance to educational researchers. Nonetheless, it is not just the need to learn complex statistical techniques that presents an issue to educational psychologists, but also the need for informed judgment as to *when* such techniques are warranted. Complex statistical analysis should be used judiciously and thoughtfully and not at the exclusion of either qualitative and mixed-method approaches. Further, these complex quantitative techniques need to be communicated in a manner that does not alienate practitioners and policymakers who possess considerably less knowledge of such methods. This particular side effect will be revisited when I forward my prognostications for educational psychology's future contributions.

Evidence-Based Approaches and Practices

Of all the contributions that need to be brought to the forefront in this treatise and one that, in many ways, is the culmination of all the other contributions documented herein, is the legacy of instructional approaches and practices shown to improve the learning and performance of students, to enhance the learning environment, or to address concomitant social/emotional concerns. These evidence-based approaches and practices have withstood the trials and tribulations that inevitably come from engaging in research within dynamic and often unpredictable contexts in schools and in the world at large (Mayer & Alexander, 2017). In her introduction to a special issue on school-based interventions, Murphy (2015) expressed the following:

> The challenge is that conducting empirically based research in classroom settings is, at best, difficult. Indeed, as is evident in the pages of this special issue, navigating the dynamic complexities of classrooms, situating relevant interventions within existing curricula, dealing with varying student abilities, school cultures, classroom enclaves, pedagogical nuances, and a general malaise toward research is no simple task. (p. 1)

Murphy's perspective on these challenges comes from middle-school and secondary-school projects on quality talk in language arts and science classrooms that she has been conducting with colleagues like Jeffrey Greene (Murphy, Firetto, & Greene, 2016; Murphy, Firetto, Wei, Li, & Croninger, 2016). Of course, there are others within the educational psychology community who have devoted themselves to devising effective classroom-based interventions that serve identified populations who might otherwise struggle. Harris and Graham, as a case in point, have labored for more than 30 years to hone their highly regarded self-regulation strategy development (SRSD) model that has been shown to facilitate strategic behaviors, self-regulation skills, content knowledge, and motivation for writing for students with and without learning disabilities (Graham & Harris, 2016; Graham, Harris, &

Mason, 2005; Harris & Graham, 1985; Harris, Graham, & Mason, 2006).

What seems characteristic of these successful school-based projects is that they represent collaborative partnerships in which school personnel *and* university researchers jointly set out to address a concern of importance to both parties. The insights and experiences of practicing educators matter in the formulation of goals and in the development of materials. It would seem that current educational psychologists, with their investment in school-based and classroom-based research, are far removed from the days of laboratory studies of pigeons and mice. Indeed, these efforts to trial approaches and practices within ecologically valid settings stand in sharp contrast to the admonitions of E. L. Thorndike (1910) to avoid the messiness of schools and classrooms in favor of the more controllable laboratory study (Berliner, 2006).

Regardless of the obstacles and challenges they face, educational psychology researchers have for generations devised techniques, crafted measures, and designed interventions that function for certain learners of certain ages undertaking certain academic tasks (e.g., Cronbach & Snow, 1977; Goska & Ackerman, 1996). For instance, cognitive-behavioral techniques, which first appeared in the 1970s, are still considered effective tools within certain areas of practice (e.g., special education or clinical psychology) and for addressing particular behavioral or psychotherapeutic issues (Meichenbaum, 1977). Conversely, it is equally valuable that the empiricism characteristic of the educational psychology community has demonstrated the fallacy of commonly accepted beliefs or practices that would otherwise operate unabated. The investment of schools in learning styles approaches in the 1980s and 1990s (e.g., Pashler, McDaniel, Rohrer, & Bjork, 2008) or the recent push for brain-based training are two examples of initiatives that have come under scrutiny by rigorous psychological research (e.g., Ansari & Coch, 2006).

The struggle for researchers is not only to conduct investigations that consider concerns of relevance to students, teachers, administrators, and policymakers, and to do so under conditions that are ecologically valid, but also to communicate the influential outcomes to decision makers. In effect, without the "intermediary inventive" minds of which James (1899) spoke or the "middle-magazines" that the first editors of *Journal of Educational Psychology* (Bagley et al., 1910) sought to be, the Herculean labors of educational psychological researchers may lay fallow. Rather than placing the burden of communication and translation squarely on the backs of individual scholars or even specialty publications including research briefs or practitioner-oriented journals, alternatives need to be considered.

One such alternative is the creation of the What Works Clearinghouse (WWC). Established in 2002 by the United States Department of Education's Institute of Education Sciences (IES), the purpose of this repository is to archive those techniques or approaches that have been empirically shown to produce positive outcomes in certain populations. As the IES describes it:

> The What Works Clearinghouse (WWC) reviews the existing research on different programs, products, practices, and policies in education. Our goal is to provide educators with the information they need to make evidence-based decisions. We focus on the results from high-quality research to answer the question "What works in education?" (Institute of Education Sciences [IES], 2016)

For example, assume a teacher is trying to locate strategies for teaching her elementary students how to write more effectively. Logging into the WWC site, this teacher can secure a useful practice guide for teaching elementary students how to become effective writers (Graham et al., 2012) that builds on the decades of SRSD research by Harris and Graham (e.g., Graham et al., 2005; Harris et al., 2006) described earlier. The WWC is not without its critics, of course, who view the standards of evidence required for inclusion to be unduly restrictive or potentially misleading—overlooking alternative sources of empirical evidence than experimental, quasi-experimental, or regression discontinuity designs (Schoenfeld, 2006).

While referencing the WWC as one medium for research communication, I do not want to leave the impression that only studies conducted within schools or classrooms meet the criteria of evidence-based or that only experiments count as quality research. Much can be learned about effective techniques or practices from quality research conducted in laboratories, online, in workplaces, or in homes—anywhere where learning and performance can be documented. There is also something to be learned from studies that are not effective, when potential barriers to effect can be ascertained (Murphy, 2015). Of course, it should not be assumed that findings derived from projects not conducted in classrooms could be directly or easily transferred into school settings—if that is the ultimate goal. In that instance, there would need to be intermediate steps taken to ensure successful migration.

So, what are the consequences that merit explication with regard to this major contribution of the educational psychology community? Several of those consequences have already been suggested. For instance, in addition to the uncertain and messy nature of school-based and classroom-based research, there are the incredible resource demands on researchers and educational systems. Securing external funds to allay those resource demands is one option. Yet, many valuable research projects do not fit neatly into the funding agendas from leading federal agencies or foundations. In addition, there is the reality that some within the educational psychology community who are especially skilled at designing and conducting significant research may not be equally skilled at promoting those findings to nonresearch audiences.

I concur with Berliner (2006) that the "middle-man" role may not be sufficient to ensure a solid relationship between educational psychologists and educational practitioners. However, many within the educational psychology community have only a weak understanding of what transpires in the lives of teachers and students. Their earlier professional years were not spent in classrooms but in research labs or university classrooms. Thus, there does seem to be the need to identify those within the community who have one foot in the world of theory and research and another in either the world of school-based or classroom-based learning. Such dual expertise would seem invaluable for the translation and implementation of evidenced-based approaches and practices in effective educational practices.

Further, even when effective communication efforts are undertaken, those critical findings may be overlooked or discounted by policymakers. In effect, even when significant and potentially transformational outcomes arise from empirical inquiry, those

outcomes may not be sufficiently reflected in educational policies and practices. Liaisons within professional organizations, institutions and societies may prove helpful in this regard but there are no assurances that policies especially at the state or national levels will reflect best practices as delineated by empirical research. However, educational psychology researchers should not be discouraged and must persist at getting their work noticed and their empirically tested approaches and techniques woven into the fabric of instructional practice.

The Progenies

Before I share my vision for educational psychology's future contributions to schools and society, let me set forth certain framing principles that gave rise to my specific prognostications. For one, there is good reason to assume that the core contributions that have defined the educational psychology community for the past 125 years will remain core to the field's identity in the next quarter century. There is no reason to predict otherwise. In essence, psychologizing, interdisciplinarity, learning, variability and assessment, and a search for effective approaches and practices will remain evident. However, what *will* change dramatically are conditions within the world, in the educational experience, and in learners. It will be those conditions that will significantly determine how educational psychology's core contributions are ultimately embodied and enacted. Interestingly, the conditions to which I will refer already exist. In the ensuing years, however, their presence and influence will grow exponentially. Thus, it will be for members of the educational psychology community to prepare themselves for the transformations that appear on the horizon.

The Conditions of Change

In terms of the external forces that will greatly impinge on the discipline of educational psychology and its abiding concern from learning and performance, there are three interrelated phenomena that must be acknowledged. The first is the relentless and unceasing production of information that occurs moment by moment across the globe. As has been proclaimed: data never sleep. In fact, according to one business information management firm, within the span of one minute:

- YouTube users upload 48 hours of new video,
- Instagram users share 3,600 new photos,
- brands and organizations on Facebook receive 34,722 "likes," and
- over 100,000 tweets are sent (DOMO, 2016).

What the aforementioned examples also amplify is that this deluge of information takes many forms—pictorial, graphic, linguistic, and auditory. Also, this multimodal, multimedia information is being transmitted at increasingly faster speeds and through social media channels that did not exist 5 or 10 years earlier (Chen, Pedersen, & Murphy, 2011).

Alongside this flood of information is the fact that technological hardware and software are also advancing in leaps and bounds (O'Neil & Perez, 2013). One aspect of those technological advancements occurring now is in the realm of virtual and augmented or mediated reality programming and devices. For instance, there is already a range of headsets available on the commercial market that permit users to experience vivid alternative realities in 3D. The popularity of these virtual worlds has been evident for some time, like Second Life, which has countless devotees who seek out the opportunity to move from the real to the virtual. In addition to virtual reality, there are augmented realities that have one foot in the real world and one in the virtual world. As with the PokémonGo, a current craze, computer-generated sensory input (e.g., images, videos, or sound) is projected onto the real, physical surroundings. Although these virtual and augmented worlds have mostly been used in gaming, there is every reason to assume that their continued advancements will have much broader applicability in the years to come (de Jong et al., 2013; Seidel & Chatelier, 2013).

Finally, in terms of technological hardware, what once required room-size equipment to operate can now be carried on one's wrist or attached to eyewear. In 2016, Jean-Pierre Sauvage, Sir James Fraser Stoddart, and Bernard Feringa were awarded the Nobel Prize in chemistry for their work on molecular machines: specially designed molecules the width of human hair. These molecular machines come with removable components that could eventually be used as energy-storing devices in the human body. Certainly, nanotechnology already exists that would ultimately permit devices of information transmission to be located internally rather than externally (Brayner, Fiévet, & Coradin, 2013). In effect, there would be no need to externally access computer technology because the device would literally be embodied. With the addition of speech interpretation and recognition interfaces (e.g., Siri) that individuals now activate on their "smart" devices, these internal technologies could also be voice activated (if not eventually thought activated).

These conditions of change are not science fiction. They already exist and operate in the world. They just have not become widespread in individuals' lives or in their education, or highly influential in educational psychology research. What matters for this discussion is the way in which these conditions will shape educational psychology's future contributions, not just in terms of the research questions posed or theoretical models derived, but also in terms of the methods and techniques created and validated. Another reason why this miniaturization and internationalization of technology matters for the future may not be readily apparent. Since the dawn of time each new technological advancement—from chalkboards to smartboards—has engendered debates as to its potential productive and destructive qualities (Cuban, 2001; Postman, 1992). Yet, when it comes to the technologies of today and tomorrow, there is one significant difference that must be acknowledged. What we now witness is that today's digital natives are almost never separated from their technologies—they hold them in their hands, wear them on their wrist, or carry them in their pockets. That means there may be limited opportunity or desire among the populace to disconnect from their smart technologies. That degree of inseparability has never occurred with past technological advancements and significantly raises the potential for influence.

With such dramatic changes looming on the horizon, new philosophical quandaries will also manifest. Specifically, while there is little doubt that the community will continue to grapple with fundamental epistemic issues, such as what constitutes viable evidence and what distinguishes knowledge from information (Alexander, 2016b; Alexander, Winters, Loughlin, & Grossnickle,

2012), it is quite likely that even more concern will be directed toward questions of ontology (Perkins, Jay, & Tishman, 1993). That is to say, educational psychologists will need to ponder questions of the reality or essence of constructs and entities that have been taken for granted. For instance, there may well be the pressing need to reconsider the very nature of learning, to reevaluate what qualifies as learning environments, or to ascertain what can be classified as an educational experience.

With this preface in mind, let me now turn to what I regard as a possible future for educational psychologists and the ensuing contributions that educational psychologists will make to schools and society in the next quarter century. Because I am moving into the realm of speculative psychology, these prognostications will be relatively brief in description and unburdened by empirical evidence. In keeping with the precedent I set in the discussion of legacies, however, I will examine these future contributions with an eye toward serious concerns that they may generate.

Paralleling the five themes of past contributions, the five predictions I proffer are:

- delving more deeply into the mind-in-context and context-in-mind,
- contemplating what it means to learn with and without technological enhancements,
- rethinking the parameters of individual variability,
- reconsidering of the form and function of assessment, and
- crafting individualized interventions for all students.

Mind-in-Context and Context-in-Mind

Over the course of the past generation, educational psychologists have become increasingly attuned to the significance of the "where" of human learning (Alexander, Schallert, & Reynolds, 2009). As many have concluded, learning is inevitably situated, and the features or affordances of that context matter greatly to the cognitive, social, physiological, motivational, and affective outcomes manifested (Westera, 2011). However, in the past, the notion of context has referred to the external environment in which learning and performance take place or to computer-generated situations (i.e., mind-in-context).

Yet, there has always been another context that has operated, typically locked away from view—the context that exists within the mind of the learner (i.e., context-in-mind). In the future, this internal context may become accessible in ways never conceived before. Learners' thoughts, actions, and emotional responses may become more directly measureable. Further, the advancements in virtual technological hardware and software may result in the creation of new internal worlds that come with their own realities. These advancements may also allow for the untethering of the educational experience from traditional physical contexts more than ever before and even from the external realities that exist outside of schools and classrooms. Overall, such events would markedly alter current notions of ecological validity.

By drawing this distinction between mind-in-context and context-in-mind, I do not want to leave the impression that these are isolated and entirely separable phenomena. To the contrary, there will inevitably be interplay between these two critical contexts. Further, the health and well-being of individuals, to say nothing of their learning and performance, demand the negotiation between both external and internal contexts. Moreover, members of the educational psychology community will continue to delve into that interface between these two "realities," and the consequences for learners and teachers alike. Nonetheless, the paths of inquiry pursued in the future and, thus, resulting contributions, will be shaped not only by the growing presence of virtual and alternative realities. They will also be altered by the deluge and variability of information that exists in that world—information readily accessed by technology.

Navigating limitless sources and questionable content. There is ample evidence in the extant literature that students are not particularly facile at choosing credible sources when navigating the Internet (Bråten, Strømsø, & Salmerón, 2011; List, Grossnickle, & Alexander, 206). Once inside those sources, students struggle to distinguish between important versus trivial, relevant versus irrelevant, and reliable versus unreliable content. These conditions will become even more of a factor when the expanding informational flood ensures an epidemic of questionable content (Acemoglu, Ozdaglar, & ParandehGheibi, 2010; Del Vicario et al., 2016). How do students deal with this mélange of data and how can they be guided to make intelligent and justified decisions about the sources they select and the content to which they attend?

Such core epistemic issues will occupy the community of educational psychologists for years to come. The current models of multiple source use will assist in these explorations (e.g., Rouet & Britt, 2011), but it is also conceivable that existing models will require refinement and renegotiation as the pressures of informational navigation and the multitude of source types increase exponentially. What this discussion suggests is that, in the future, more research efforts and resulting contributions must, by necessity, center on the critical analysis of information, as well as on the formation of reasoned decisions as to what counts as credible sources, reliable content, and viable evidence. Granted, this call for more critical analysis and reasoned decisions is by no means new, and the results of ongoing efforts to hone evidenced-based decisions may not engender great optimism. Nonetheless, in a world already plagued by "fake news" and "alternative facts," it is even more imperative that an informed segment of the world's population can cull the wheat from the chaff when it comes to the information onslaught.

Information management versus knowledge building. Recently, in confronting conditions within schools and society that involve the increase in information in the external environment coupled with individuals' fascination with technology and schools' obsession with assessment, I have distinguished between information management and knowledge building to bring learners' intentions to bear on the mind-in-context situation (Alexander, 2016b, in press). In effect, the contention is that individuals may adopt the mindset of gathering whatever information seems workable for some particular task without being overly evaluative and with no goal of retaining the gathered information beyond task performance (i.e., information management). In contrast, there are those exploring a topic of interest, addressing a personally relevant question, or tackling a problem of value. For these individuals, information management is insufficient. They must be more thoughtful in the data they extract from the informational universe, and they must then critically examine that information and formulate some representation that can be retained in memory. In effect, these individuals must be effective knowledge builders.

Although effective learning and performance undoubtedly involve both of these orientations toward information, what needs to be better understood is *when*, *why*, and *how* individuals engage in information management and knowledge building. Simply due to the rise in information encountered, the challenges of living and learning in the world of tomorrow could potentially exacerbate the occasions of information management and depress motivations for knowledge building. Thus, in the future, educational psychologists' contributions will entail the richer characterization of effective information management and knowledge building, as well as deal with more complex questions about the conditions under which one or the other of these orientations should hold sway in the learning environment. Further, making students more cognizant of their intentions and the consequences of those intentions on the processing that transpires would seem a valuable step in crafting learning environments and educational tasks that elicit knowledge building.

Significance of metastrategic processes. Collectively, these mind-in-context examples suggest that educational psychologists will need to contribute further to current understandings about the role of metastrategic processes, such as metacognition, self-regulation, and relational reasoning, in living and learning in an information-saturated world (Winters, Greene, & Costich, 2008). It is conceivable that the lack of effective monitoring or regulating of thinking and performing will make it more likely that learners lose themselves in the informational deluge. Further, without the capacity to perceive meaningful patterns within the onslaught of seemingly unrelated information (i.e., relational reason ability), those learners will be hampered in the efforts to make sense of the dynamic world that surrounds them (Dumas, Alexander, & Grossnickle, 2013). The question for the community of educational psychology is how to help in this process. What more can be done to identify and promote metastrategies within the educational experience that will undoubtedly involve online elements? These are among the contributions that educational psychologists will need to even more aggressively pursue in the next quarter century.

Virtual or augmented realities versus real realities. These invaluable lines of inquiry related to how learners interface with the external environment and what that environment affords is only half of the picture. As I have suggested, there is also the likelihood that the students of tomorrow will be able to enact very vivid, alternative realities that are loosely tethered to their real surroundings, if at all. With the combination of nanotechnologies and the rapid growth of virtual and augmented reality programs, rich contexts existing solely or partially within the mind of the learner are conceivable. Educational psychologists will need to be in the vanguard of those who attempt to ascertain the likely promises and pitfalls of such context-in-mind situations (Brayner et al., 2013). It is at this juncture that basic concerns about the very nature of a learning environment will require interrogation, along with questions about the place of virtual or augmented realities versus real realities on what is taught and what is learned.

It is certainly plausible that the advancements in virtual and augmented reality technologies could prove invaluable in crafting simulations that promote deeper learning and active engagement in complex and dynamic tasks, including tasks that students would be unlikely to experience in the real world (de Jong et al., 2013). These vivid contexts in the mind could also be put to good use when individuals need to be trained in certain procedures that would be too expensive or too risky to attempt in reality (de Jong, 2011). Perhaps one of the most intriguing applications of these virtual technologies would be in the development of virtual courses or even virtual classrooms that could replace the online or hybrid delivery systems that currently operate. These virtual courses and classrooms could permit those who cannot physically attend or emotionally negotiate real classrooms with all their cognitive-social dynamism to attend virtually (Lorenzo, Pomares, & Lledó, 2013). Nonetheless, serious concerns remain for these context-in-mind scenarios, including increased distractibility, dissociative behaviors, and the challenge to distinguish between the real and imagined, as well as the "authentic" versus contrived (Loh & Kanai, 2016; Weinstein & Lejoyeux, 2010).

Learning With and Learning Without Technological Enhancements

In the preceding discussion, there were serious questions about the nature of contexts that will mark learning and performance in the future. What these contextual elements will undoubtedly bring to the surface are serious ontological questions about the nature of learning. These will surface not just because of the intriguing contrasts between in-the-environment and in-the-mind contexts or because of the increasing pressures to keep one's head above water in an information-saturated world. There are also particular affordances nested in these varied educational contexts that have the potential to enhance (and, thus, alter) the process of learning in significant ways. One can only imagine the future applications of nanotechnology and molecular machinery for human learning and performance, including as "mind enhancers" that interface with cognitive-neurological processing (Schneider, 2008).

For instance, within the literature on executive functioning, there are two capacities that are frequently described as foundational to human learning and performance—working memory and inhibitory control (Baggetta & Alexander, 2016; Meltzer, 2011). It is not far-fetched to assume that both of these capacities could be dramatically transformed, for better or worse, through nanotechnologies. For example, just as one can augment the memory of external technological devices, it is conceivable that individuals could rely on an internal memory chip to enhance or complement their memory (Sandberg & Bostrom, 2006). Similarly, it is conceivable that internal devices could be programmed to serve as monitoring or regulatory prompts to keep individuals more focused, attentive, and on task or as cognitive stimulators to jolt or arouse neuro-functioning (Heller & Peterson, 2007). These possibilities raise questions about whether the learning that transpires under these internally enhanced conditions would be appreciably better or markedly different than what would occur without such enhancements.

On the one hand, there is the prospect that freeing up memory demands or prompting monitoring or regulation could assist learners with executive functioning issues. With these internal enhancements, those who have been challenged in their learning processes might actually meet or surpass those without such challenges—narrowing the performance gap. On the other hand, because these internal enhancements would potentially be available to all learners, and not just those with some demonstrated needs, the conse-

quences might well be reversed. That is, what might result, in keeping with the principle of the Matthew effect (Stanovich, 1986), is that the existence of cognitive-neurological enhancements could exacerbate differences, creating even wider disparities among student populations.

What is the evidence for a claim that exacerbated differences among individuals could emerge if cognitive-neurological enhancements became available? Certainly, the research on computer usage is informative on this point. There are already many ways in which existing technologies have been used to augment performance, including through computer-assisted and computer-adaptive programming (Mathes, Torgesen, & Allor, 2001; Scheiter & Gerjets, 2007). Conversations around the "digital divide" in the early 2000s suggested that the lack of access to computer technologies was a negative force for students from minority populations. Without such access, it was argued, minority students would fall farther behind their more technologically savvy peers. Yet, contemporarily, and across the socioeconomic spectrum, it is rare to find students without access to the Internet via "smart" technologies. So computer access per se is much less of an issue. Yet, much of students' time online has nothing to do with enhanced learning. As is true for virtual reality programs and devices today, these technologies are more commonly used for recreational purposes such as gaming (Goldfarb & Prince, 2008).

Thus, it would seem that it is not simply access that matters but the *smarter use* of those smart technologies or any technologies for that matter (van Deursen & Van Dijk, 2014). What is important to acknowledge is that no technology is inherently good or bad for learning. It is the application of those technologies that likely proves a boon or a barrier to learning. In this regard, I am reminded of the wisdom of Alfred North Whitehead (1929/1967) who proclaimed that "the best education is to be found in gaining the utmost information from the simplest apparatus" (p. 11). Smarter use of smart technologies should embrace the goal of utmost information, as well.

In projecting forward, there is reason to believe that, as the next generation of technologies find its way into the lives of students, an alternative form of the digital divide related to smart use will exacerbate learner differences for better or worse. Thus, it will fall to educational psychologists, working with experts from the fields of neuroscience, special education, nanotechnology engineering, and more, to map the effects of these advancements on learning and performance and prepare viable responses that work for all students.

Parameters of Individual Variability

This discussion about the potential ramifications of new nanotechnologies, along with the expectation of living and learning in an information-saturated world, casts new light on the topic of individual differences. As described in the overview of past contributions, the concern for human variability and the concomitant study of individual differences have long been pillars of the educational psychology community (Galton, 1869; Jonassen & Grabowski, 2012; Sternberg & Grigorenko, 2003). What must now be considered is how shifting circumstances within the world and inside schools and classrooms might spur attention to certain salient differences that have, to date, gone unnoticed.

As a case in point, while attention problems and their potential effects on student learning have concerned educators for a century, it was in the 1960s that the label of "learning disabled" was first introduced. Over the course of the next 40 years, the label for those manifesting particular learning difficulties underwent change. For one, a new category of special needs, ADD (hyperkinetic impulse disorder), was introduced and then later renamed ADHD (attention-deficit hyperactivity disorder). This renaming reflected more than a name change, however. It marked a deeper understanding of the neuropsychological components associated with ADHD, such as behavioral inhibition, working memory, regulation of motivation, and motor control (Barkley, 1997). Among the controversies surrounding this category of individual differences are the over diagnosis of this condition; the over medication of those identified; and, the argument that this is a problem that is reflective of the nature of schooling in modern societies (SLeFever, Dawson, & Morrow, 1999; Singh, 2008).

I specifically bring up the subject of learning and attentional problems because it is quite likely that the conditions that give rise to this diagnosis will not abate in the next quarter century. Rather, the inundation of information, the technological advancements in hardware and software, suggest even more opportunities for diffused attention, distractibility, and dissociation. There is also a growing interest in an obsession with social media being reported, what Rosen (2012) has labeled "iDisorders." That is, individuals' compulsions to stay continually connected to others via social media come to interfere with their ability to live and learn in the "now" or to stay focused on interactions unfolding in their immediate physical environment in real time. Even among those who do not display such an obsession, the intrusion of the Internet and social media into everyday lives is evident (LaRose, Eastin, & Gregg, 2001). For example, there is a sense that individuals are required to deal with all Internet communications with urgency rather than to set them aside until time and context allow for response. I bring up these points because there is good reason to assume that there are new categories of individual differences on the horizon that reflect these changing conditions.

For example, consider the demands that the continual interface between the world in the mind and in the environment places on even normally functioning children and adults. There are already identified conditions that arise when these two "contexts" cannot be effectively negotiated, resulting in social-communicative disorders or dissociation. With the increased complexity and vividness of these internal and external contexts, it is reasonable to assume that new categories of maladaptive behavior will be recognized. When this occurs, it will fall to the educational psychology community to bring its expertise to bear. Understanding the nature of these conditions, the contributory factors, the influence on human learning and performance, and effective interventions will be among the important contributions that educational psychologists make to schools and society.

Form and Function of Assessment

Educational psychology's investment in assessment will continue undaunted into the next quarter century. However, there will likely be significant transformations in the form of those assessments and their function that will be evidenced as a consequence of the conditions for change previously outlined. For example,

anyone who has attempted to gather real-time processing data or neuroimaging data can appreciate how time-consuming and expensive that research can be (King, 2011; Poline et al., 2012). In addition, the settings under which such data are collected generally bear little resemblance to "authentic" conditions under which individuals live and learn (Ansari & Coch, 2006). Now, imagine the possibilities for data gathering when the required data collection equipment is internal as a result of nanotechnology. In that instance, real time would take on an entirely new meaning, given the continuous and massive stream of data that would emerge for each individual. Also, those data could encompass any range of cognitive, affective, and physiological information. In this way, paper-and-pencil forms of assessment and mechanisms for data gathering that have been commonplace in the past would become increasingly obsolete and psychometrically questionable.

Further, with this marked transformation in the form and wealth of data that could be gathered directly from individuals as they engage in relevant activities comes untold possibilities for the uses to which those data could be put. At this point, it is hard to conceptualize those uses. However, I have already suggested that one possibility is to serve as an internal monitor or regulatory prompt that signals individuals when they are distracted or hyperactive or when their affective state seems off kilter (Staggers, McCasky, Brazelton, & Kennedy, 2008). Just like the activity bands that many wear to show how many steps they have taken or miles they have run, I can imagine the popularity of devices that monitor thinking and behavior and that remind users when they are cognitively idle or off task.

Of course, these transformations will not occur on their own. It would fall to members of the educational psychology community, perhaps working collaboratively with experts from biotechnology and biostatistics, to devise new procedures to manage the ocean of data that might be collected and to formulate systems that could be used to detect meaningful patterns within those data streams. Much as contemporary companies like Google or Amazon apply complex algorithms to detect user preferences and bowser patterns, tomorrow's measurement and assessment experts will need to develop procedures for handling such an enormous cache of data.

This envisioned future for assessment brings several emerging trends within the statistics and measurement domain to the forefront. One pertains to the necessary reductionism that massive amounts of data demand (Kievit et al., 2011; Stapleton, 2013). Another relates to the growing popularity of intraindividual variation models over latent variable models of data analysis (Molenaar, 2007; Molenaar, Lerner, & Newell, 2013). In effect, there is every reason to assume that the wealth of individual data to emerge in the future will require even more attention to intraindividual variations. Further, the struggles of dealing with multidimensionality witnessed in contemporary data analyses will likely expand in ensuing decades. Current work in ecological momentary assessment may prove useful in devising new statistical procedures in the decades to come (Dunton et al., 2014; Hedeker, Mermelstein, Berbaum, & Campbell, 2009).

As the prior discussion suggests, it is possible to conceive of revolutionary changes in the way learning and performance data are amassed in the future. It is also possible to appreciate the new contributions in measurement and statistical techniques that would be required to keep pace with those changes. Nonetheless, there may well be less desirable outcomes to arise. For instance, I mentioned the challenges that already exist for those engaged in the empirical analysis of learning and performance data and the level of methodological and substantive knowledge demanded of community members. I am left wondering how the community will prepare the next generation of scholars capable of acquiring the extensive methodological and substantive knowledge required. Is this a case where interdisciplinary and cross-disciplinary inquiry will become commonplace, ensuring that both the methodological and substantive knowledge are sufficiently represented?

Beyond these methodological and statistical concerns, there are complications to be weighed. For instance, consider how the prevalence of social media currently has contributed to what some would regard as a deterioration of privacy (Joinson, Houghton, Vasalou, & Marder, 2011). Lives seem open to public display. Would there be mechanisms built into these internal systems that would block certain thoughts and feelings from external probe? Who would be permitted access to such data and for what purposes? What would be the ramifications for researchers attempting to delve into those once private forms of information? What would be the safeguards to the participants whose internal thoughts and feelings were being probed or analyzed? How would the anonymity of individuals be protected? If there were a discrepancy between the internal data and individuals' verbalizations or actions, which would be considered more credible? In effect, what ethical issues should researchers of tomorrow anticipate and address when private thoughts and actions, as well as physiological and affective responses, become externalized?

Individualized Interventions

When considering educational psychology's past contributions, I concluded by lauding the community's efforts to identify approaches and practices that work to promote learning and performance. In considering future contributions, I will do so again. Specifically, by drawing on the advancements outlined in this examination of the community's tomorrows—alternative contexts, enhanced learning, expanded notions of individual differences, and new forms of and functions for assessment—educational psychologists committed to uncovering empirically supported approaches and practices will have options not previously available to them.

One clear option, which is already being pursued within medical professions, is the opportunity to devise treatments and interventions tailored to the needs of the individual learner and not to some specified class or group of learners (Parveen, Misra, & Sahoo, 2012). This new trend has been labeled precision medicine (National Research Council Committee on A Framework for Developing a New Taxonomy of Disease, 2011). Using precision medicine as a model, educational psychology researchers would be involved in the formulation of "boutique" treatments that effectively address a specific pattern of thought, behavior, or social-emotional response in a learner in lieu of seeking out tested approaches for entire classes or categories of learners. Despite the incredible complexity that this individuation would demand, there are broad implications that might be realized. For one, it would be possible for the education system to be less focused on learner deficits. These individualistic treatment models could just as well target identified strengths in learner profiles than points of weakness. These areas of strength in these individual models could be used as leverage to address demonstrated needs. Moreover, the

existence of these individualistic treatments built on very learner-specific data may lessen the emphasis on high-stakes assessments or the need to compare classes of students in terms of their achievement.

I can certainly see how many of the existing interventions that have proven viable could be modified and individualized to work for all students. The Quality Talk research (Li et al., 2016; Murphy, Firetto, et al., 2016; Murphy, Firetto, Wei, et al., 2016), for instance, could be augmented via nanotechnology to allow individual students to engage in rich and evidence-based discussions internally with a virtual peer whose level of knowledge and facility at argumentation was expressly matched to the level of that student. How would the ecological validity of these internal discussions be judged? Further, the capacity of the virtual discussion peer could be programmed to modulate as the demands of the task or text shift or as the capabilities of the learner to pose "authentic" questions and engage in evidenced-based discussion develops.

Similarly, some of the external cues and prompts that have been effectively employed within the learning environment for SRSD (Graham et al., 2012; Graham et al., 2005; Harris et al., 2006) could be transferred to internal prompts that are automatically activated as the student engages in writing tasks. Further, those regulatory prompts could be systematically faded as the need for scaffolding diminishes. Granted, these examples are presently "pie in the sky" speculation, but the possibilities for individualization of effective interventions do exist. Moreover, it will be for the educational psychology community not only to anticipate the need for individualization and to consider what possible approaches and practices to pursue. It would also fall to members of the educational psychology community to design learning environments that reduce unnecessary sources of variance and that seek to balance individualization with shared learning experiences. In addition, it will be a significant contribution of educational psychologists to put individual treatments and interventions to rigorous testing to ensure that they do, in fact, work, whether at the level of the individual student or entire populations of learners. This empirical validation is what educational psychologists have always done and will, undoubtedly, continue to do for decades to come.

Concluding Thought

It has been my intention in this treatise to bring the many significant contributions of educational psychology over the past 125 years to light. Those contributions are the very milestones that have marked our journey as community members and that reveal how far the discipline has progressed in even its relatively short history. Using those past milestones as the means of reckoning, I have also attempted to project forward—to envision the contributions that lie ahead. Only time will tell whether there is any truth to my speculations. However, if my ramblings cause reflection, instigate debate, or spark response, then this effort has been worthwhile.

References

Acemoglu, D., Ozdaglar, A., & ParandehGheibi, A. (2010). Spread of (mis) information in social networks. *Games and Economic Behavior, 70,* 194–227. http://dx.doi.org/10.1016/j.geb.2010.01.005

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32,* 10–14. http://dx.doi.org/10.3102/0013189X032008010

Alexander, P. A. (2016a). The arguments for and the reasoning about epistemic cognition: A response to the chapters on psychological perspectives. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 100–110). New York, NY: Routledge.

Alexander, P. A. (2016b) *Re-educating the mind* [Video file]. Retrieved from https://www.youtube.com/watch?v=ozIapr_MmU8

Alexander, P. A. (in press). Information management versus knowledge building: Implications for learning and assessment in higher education. In O. Zlatkin-Troitschanskaia & H. A. Pant (Eds.), *Assessment of learning outcomes in higher education: Cross national comparisons and perspectives.* Dordrecht, the Netherlands: Springer.

Alexander, P. A., & Murphy, P. K. (1998). The research base for APA's learner-centered principles. In N. M. Lambert & B. L. McCombs (Eds.), *Issues in school reform: A sampler of psychological perspectives on learner-centered schools* (pp. 25–60). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10258-001

Alexander, P. A., Murphy, P. K., & Greene, J. A. (2012). Projecting educational psychology's future from its past and present: A trend analysis. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *Educational psychology handbook* (Vol. 1, pp. 3–32). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/13273-001

Alexander, P. A., & Riconscente, M. M. (2005). A matter of proof: Why achievement ≠ learning. In J. S. Carlson & J. R. Levin (Eds.), *The No Child Left Behind legislation: Educational research and federal funding* (pp. 27–36). Greenwich, CT: Information Age.

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research, 61,* 315–343. http://dx.doi.org/10.3102/00346543061003315

Alexander, P. A., Schallert, D. L., & Reynolds, R. E. (2009). What is learning anyway? A topographical perspective considered. *Educational Psychologist, 44,* 209–214. http://dx.doi.org/10.1080/00461520903029006

Alexander, P. A., Winters, F. I., Loughlin, S. M., & Grossnickle, E. M. (2012). Students' conceptions of knowledge, information, and truth. *Learning and Instruction, 22,* 1–15. http://dx.doi.org/10.1016/j.learninstruc.2011.05.004

Ansari, D., & Coch, D. (2006). Bridges over troubled waters: Education and cognitive neuroscience. *Trends in Cognitive Sciences, 10,* 146–151. http://dx.doi.org/10.1016/j.tics.2006.02.007

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and operationalization of executive function. *Mind, Brain, and Education, 10,* 10–33. http://dx.doi.org/10.1111/mbe.12100

Bagley, W. G., Bell, J. C., Seashore, C. E., & Whipple, G. M. (1910). Editorial. *Journal of Educational Psychology, 1,* 1–3. http://dx.doi.org/10.1037/h0065787

Barkley, K. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121,* 65–94.

Bell, J. C. (1914). Editorial: Retrospect and Prospect. *Journal of Educational Psychology, 1,* 42–43.

Berliner, D. C. (2006). Educational psychology: Searching for essence throughout a century of influence. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 3–27). New York, NY: Taylor & Francis.

Berliner, D. C. (2008). Research, policy, and practice: The great disconnect. In S. D. Lapan & M. T. Quartaroli (Eds.), *Research essentials: An introduction to designs and practices* (pp. 295–325). Hoboken, NJ: Jossey-Bass.

Berliner, D. C. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal*

of Education, 41, 287–302. http://dx.doi.org/10.1080/0305764X.2011.607151

Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction, 21,* 180–192. http://dx.doi.org/10.1016/j.learninstruc.2010.02.002

Brayner, R., Fiévet, F., & Coradin, T. (Eds.). (2013). *Nanomaterials: A Danger or a Promise?* London, UK: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4471-4213-3

Bredo, E. (2003). The development of Dewey's psychology. In B. J. Zimmerman & D. H. Schunk (Eds.), *Educational psychology: A century of contributions* (pp. 81–111). Mahwah, NJ: Erlbaum.

Broekkamp, H., & van Hout-Wolters, B. (2007). The gap between educational research and practice: A literature review, symposium, and questionnaire. *Educational Research and Evaluation, 13,* 203–220. http://dx.doi.org/10.1080/13803610701626127

Cervetti, G. N., Kulikowich, J. M., & Bravo, M. A. (2015). The effects of educative curriculum materials on teachers' use of instructional strategies for English language learners in science and on student learning. *Contemporary Educational Psychology, 40,* 86–98. http://dx.doi.org/10.1016/j.cedpsych.2014.10.005

Chen, C. Y., Pedersen, S., & Murphy, K. L. (2011). Learners' perceived information overload in online learning via computer-mediated communication. *Research in Learning Technology, 19,* 101–116. http://dx.doi.org/10.1080/21567069.2011.586678

Chinn, C. A., Buckland, L. A., & Samarapungavan, A. L. A. (2011). Expanding the dimensions of epistemic cognition: Arguments from philosophy and psychology. *Educational Psychologist, 46,* 141–167. http://dx.doi.org/10.1080/00461520.2011.587722

Clements, D. H., & Sarama, J. (2004). Learning trajectories in mathematics education. *Mathematical Thinking and Learning, 6,* 81–89. http://dx.doi.org/10.1207/s15327833mtl0602_1

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions.* Oxford, UK: Irvington.

Cuban, L. (2001). *Oversold and underused: Reforming schools through technology 1980–2000.* Cambridge, MA: Harvard University Press.

Darling-Hammond, L. (1992). Educational indicators and enlightened policy. *Educational Policy, 6,* 235–265. http://dx.doi.org/10.1177/0895904892006003001

DeCuir-Gunby, J. T. (2008). Mixed methods research in the social sciences. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 125–136) Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412995627.d11

de Jong, T. (2011). Instruction based on computer simulations. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 446–466). New York, NY: Routledge.

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340,* 305–308. http://dx.doi.org/10.1126/science.1230579

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., . . . Quattrociocchi, W. (2016). The spreading of misinformation online. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 113,* 554–559. http://dx.doi.org/10.1073/pnas.1517441113

Denissen, J. J., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development, 78,* 430–447. http://dx.doi.org/10.1111/j.1467-8624.2007.01007.x

Dewey, J. (1902). *The child and the curriculum.* Chicago, IL: University of Chicago Press.

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20,* 391–409. http://dx.doi.org/10.1007/s10648-008-9083-6

DOMO. (2016). *Data never sleeps.* Retrieved from https://www.domo.com/learn/infographic-data-never-sleeps

Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: A systematic review of the literature. *Educational Psychology Review, 25,* 391–427. http://dx.doi.org/10.1007/s10648-013-9224-4

Dunton, G. F., Huh, J., Leventhal, A. M., Riggs, N., Hedeker, D., Spruijt-Metz, D., & Pentz, M. A. (2014). Momentary assessment of affect, physical feeling states, and physical activity in children. *Health Psychology, 33,* 255–263. http://dx.doi.org/10.1037/a0032640

Ercikan, K., & Roth, W. M. (2009). *Generalizing from educational research: Beyond qualitative and quantitative polarization.* New York, NY: Routledge.

Every Student Succeeds Act (ESSA) of 2015. Pub. L. No. 114–95, 114 U.S.C. (2015).

Farley, F., Alexander, P. A., Baker, E. L., Berliner, D. C., Calfee, R. C., De Corte, E., & Mayer, R. (2015). Afterwords I: Perspectives on the past, present, and future of educational psychology. In L. Corno & E. Anderman (Eds.), *Handbook of educational psychology* (3rd ed., pp. 415–432). New York, NY: Routledge.

Galton, F. (1869). *Hereditary genius: An inquiry into its laws and consequences* (Vol. 27). New York, NY: Macmillan. http://dx.doi.org/10.1037/13474-000

Goldfarb, A., & Prince, J. (2008). Internet adoption and usage patterns are different: Implications for the digital divide. *Information Economics and Policy, 20,* 2–15. http://dx.doi.org/10.1016/j.infoecopol.2007.05.001

Goska, R. E., & Ackerman, P. L. (1996). An aptitude–treatment interaction approach to transfer within training. *Journal of Educational Psychology, 88,* 249–259. http://dx.doi.org/10.1037/0022-0663.88.2.249

Graham, S., Bollinger, A., Booth Olson, C., D'Aoust, C., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide.* Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Graham, S., & Harris, K. R. (2016). Evidence-based practice and writing instruction. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed., pp. 211–226). New York, NY: Guilford Press.

Graham, S., Harris, K. R., & Mason, L. H. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology, 30,* 207–241. http://dx.doi.org/10.1016/j.cedpsych.2004.08.001

Harris, K., & Graham, S. (1985). Improving learning disabled students' composition skills: Self-control strategy training. *Learning Disability Quarterly, 8,* 27–36. http://dx.doi.org/10.2307/1510905

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43,* 295–340. http://dx.doi.org/10.3102/00028312043002295

Harris, K. R., Graham, S., & Urdan, T. (Eds.). (2012). *APA educational psychology handbook: Vol. 1–3. Theories, constructs, and critical issues.* Washington, DC: American Psychological Association.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Hedeker, D., Mermelstein, R. J., Berbaum, M. L., & Campbell, R. T. (2009). Modeling mood variation associated with smoking: An application of a heterogeneous mixed-effects model for analysis of ecological momentary assessment (EMA) data. *Addiction, 104,* 297–307. http://dx.doi.org/10.1111/j.1360-0443.2008.02435.x

Heller, J., & Peterson, C. (2007). Nanotechnology: Maximizing benefits,

minimizing downsides. In N. M. de, S. Cameron & M. E. Mitchell (Eds.), *Nanoscale: Issues and perspectives for the nano century* (pp. 83–96). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9780470165874.ch7

Institute of Education Sciences (IES). (2016, November 28). *What Works Clearinghouse*. Retrieved from https://ies.ed.gov/ncee/WWC/

James, W. (1899). *Talks to teachers on psychology: And to students on some of life's ideals*. New York, NY: Henry Holt. http://dx.doi.org/10.1037/10814-000

Joinson, A. N., Houghton, D. J., Vasalou, A., & Marder, B. L. (2011). Digital crowding: Privacy, self-disclosure, and technology. In S. Trepte & L. Reinecke (Eds.), *Privacy online. Perspectives on privacy and self-disclosure in the social web* (pp. 33–46). Berlin, Germany: Springer.

Jonassen, D. H., & Grabowski, B. L. (2012). *Handbook of individual differences, learning, and instruction*. New York, NY: Routledge.

Kievit, R. A., Romeijn, J. W., Waldorp, L. J., Wicherts, J. M., Scholte, H. S., & Borsboom, D. (2011). Mind the gap: A psychometric approach to the reduction problem. *Psychological Inquiry, 22,* 67–87. http://dx.doi.org/10.1080/1047840X.2011.550181

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science, 331,* 719–721. http://dx.doi.org/10.1126/science.1197872

LaRose, R., Eastin, M. S., & Gregg, J. (2001). Reformulating the Internet paradox: Social cognitive explanations of Internet use and depression. *Journal of Online Behavior, 1,* 1–19. Retrieved from http://www.behavior.net/JOB/v1n2/paradox.html

Lee, W., Lee, M. J., & Bong, M. (2014). Testing interest and self-efficacy as predictors of academic self-regulation and achievement. *Contemporary Educational Psychology, 39,* 86–99. http://dx.doi.org/10.1016/j.cedpsych.2014.02.002

LeFever, G. B., Dawson, K. V., & Morrow, A. L. (1999). The extent of drug therapy for attention deficit-hyperactivity disorder among children in public schools. *American Journal of Public Health, 89,* 1359–1364. http://dx.doi.org/10.2105/AJPH.89.9.1359

Li, M., Murphy, P. K., Wang, J., Mason, L. H., Firetto, C. M., Wei, L., & Chung, K. S. (2016). Promoting reading comprehension and critical-analytic thinking: A comparison of three approaches with fourth and fifth graders. *Contemporary Educational Psychology, 46,* 101–115. http://dx.doi.org/10.1016/j.cedpsych.2016.05.002

Loh, K. K., & Kanai, R. (2016). How has the internet reshaped human cognition? *The Neuroscientist, 22,* 506–520. http://dx.doi.org/10.1177/1073858415595005

Lorenzo, G., Pomares, J., & Lledó, A. (2013). Inclusion of immersive virtual learning environments and visual control systems to support the learning of students with Asperger syndrome. *Computers & Education, 62,* 88–101. http://dx.doi.org/10.1016/j.compedu.2012.10.028

Marsh, H. W. (1990). Causal ordering of academic self-concept and academic achievement: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82,* 646–656. http://dx.doi.org/10.1037/0022-0663.82.4.646

Mason, L., Boscolo, P., Tornatora, M. C., & Ronconi, L. (2013). Besides knowledge: A cross-sectional study on the relations between epistemic beliefs, achievement goals, self-beliefs, and achievement in science. *Instructional Science, 41,* 49–79. http://dx.doi.org/10.1007/s11251-012-9210-0

Mathes, P. G., Torgesen, J. K., & Allor, J. H. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal, 38,* 371–410. http://dx.doi.org/10.3102/00028312038002371

Mayer, R. E., & Alexander, P. A. (2017). *Handbook of research on learning and instruction* (2nd ed.). New York, NY: Routledge.

McMaster, K. L., van den Broek, P., Espin, C. A., Pinto, V., Janda, B., Lam, E., . . . van Boekel, M. (2015). Developing a reading comprehension intervention: Translating cognitive theory to educational practice. *Contemporary Educational Psychology, 40,* 28–40. http://dx.doi.org/10.1016/j.cedpsych.2014.04.001

Meichenbaum, D. (1977). Cognitive behaviour modification. *Cognitive Behaviour Therapy, 6,* 185–192.

Meltzer, L. (2011). *Executive function in education: From theory to practice*. New York, NY: Guilford Press.

Molenaar, P. C. M. (2007). Psychological methodology will change profoundly due to the necessity to focus on intra-individual variation. *Integrative Psychological & Behavioral Science, 41,* 35–40. http://dx.doi.org/10.1007/s12124-007-9011-1

Molenaar, P. C. M., Lerner, R. M., & Newell, K. M. (2013). *Handbook of developmental systems theory and methodology*. New York, NY: Guilford Press.

Murphy, P. K. (2015). Marking the way: School-based interventions that "work". *Contemporary Educational Psychology, 40,* 1–4.

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology *Contemporary Educational Psychology, 25,* 3–53. http://dx.doi.org/10.1006/ceps.1999.1019

Murphy, P. K., Firetto, C. M., & Greene, J. A. (2016). Enriching students' scientific thinking through relational reasoning: Seeking evidence in tasks, texts, and talk. *Educational Psychology Review, 29,* 105–117. http://dx.doi.org/10.1007/s10648-016-9387-x

Murphy, P. K., Firetto, C. M., Wei, L., Li, M., & Croninger, R. M. V. (2016). What REALLY works. *Policy Insights from Behavioral and Brain Science, 3,* 27–35. http://dx.doi.org/10.1177/2372732215624215

Murphy, P. K., & Mason, L. (2006). Changing knowledge and beliefs. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 305–324). Mahwah, NJ: Lawrence Erlbaum Associates.

Nandagopal, K., & Ericsson, K. A. (2012). Enhancing students' performance in traditional education: Implications from expert performance approach and deliberate practice. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *Educational psychology handbook* (Vol. 1, pp. 257–293). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/13273-010

National Research Council Committee on A Framework for Developing a New Taxonomy of Disease. (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, DC: National Academies Press.

Newton, K. J., Leonard, J., Evans, B. R., & Eastburn, J. A. (2012). Preservice elementary teachers' mathematics content knowledge and teacher efficacy. *School Science and Mathematics, 112,* 289–299. http://dx.doi.org/10.1111/j.1949-8594.2012.00145.x

Oakes, J. (1990). Opportunities, achievement, and choice: Women and minority students in science and mathematics. *Review of Research in Education, 16,* 153–222.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71,* 137–148. http://dx.doi.org/10.1177/001440290507100201

O'Neil, H. F., & Perez, R. S. (2013). *Web-based learning: Theory, research, and practice*. New York, NY: Routledge.

Pajares, F. (2003). William James: Our father who begat us. In B. J. Zimmerman & D. H. Schunk (Eds.), *Educational psychology: A century of contributions* (pp. 41–64). Mahwah, NJ: Erlbaum.

Parveen, S., Misra, R., & Sahoo, S. K. (2012). Nanoparticles: A boon to drug delivery, therapeutics, diagnostics and imaging. *Nanomedicine; Nanotechnology, Biology, and Medicine, 8,* 147–166. http://dx.doi.org/10.1016/j.nano.2011.05.016

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9,* 105–119. http://dx.doi.org/10.1111/j.1539-6053.2009.01038.x

Peirce, C. S. (1877, November). The fixation of belief. *Popular Science Monthly, 12,* 1–15.

Pellegrino, J. W. (2004). *The evolution of educational assessment: Considering the past and imagining the future*. Princeton, NJ: Educational Testing Service, Policy Evaluation and Research Center, Policy Information Center.

Perkins, D., Jay, E., & Tishman, S. (1993). New conceptions of thinking: From ontology to education. *Educational Psychologist, 28,* 67–85. http://dx.doi.org/10.1207/s15326985ep2801_6

Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., . . . Kennedy, D. N. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics, 6,* 9. http://dx.doi.org/10.3389/fninf.2012.00009

Postman, N. (1992). *Technopoly: The surrender of culture to technology*. New York, NY: Knopf.

Rosen, L. D. (2012). *iDisorder: Understanding our obsession with technology and overcoming its hold on us*. New York, NY: St. Martin's Press.

Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Charlotte, NC: Information Age.

Rubie-Davies, C. M., Flint, A., & McDonald, L. G. (2012). Teacher beliefs, teacher characteristics, and school contextual factors: What are the relationships? *British Journal of Educational Psychology, 82,* 270–288. http://dx.doi.org/10.1111/j.2044-8279.2011.02025.x

Sandberg, A., & Bostrom, N. (2006). Converging cognitive enhancements. *Annals of the New York Academy of Sciences, 1093,* 201–227. http://dx.doi.org/10.1196/annals.1382.015

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review, 19,* 285–307. http://dx.doi.org/10.1007/s10648-007-9046-3

Schneider, S. (2008). Future minds: Transhumanism, cognitive enhancement and the nature of persons. Retrieved from http://repository.upenn.edu/neuroethics_pubs/37

Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35,* 13–21. http://dx.doi.org/10.3102/0013189X035002013

Seidel, R. J., & Chatelier, P. R. (2013). *Virtual reality, training's future? Perspectives on virtual reality and related emerging technologies*. New York, NY: Springer Science & Business Media.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29,* 4–14. http://dx.doi.org/10.3102/0013189X029007004

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57,* 1–23. http://dx.doi.org/10.17763/haer.57.1.j463w79r56455411

Singh, I. (2008). Beyond polemics: Science and ethics of ADHD. *Nature Reviews Neuroscience, 9,* 957–964. http://dx.doi.org/10.1038/nrn2514

Staggers, N., McCasky, T., Brazelton, N., & Kennedy, R. (2008). Nanotechnology: The coming revolution and its implications for consumers, clinicians, and informatics. *Nursing Outlook, 56,* 268–274. http://dx.doi.org/10.1016/j.outlook.2008.06.004

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407. http://dx.doi.org/10.1598/RRQ.21.4.1

Stapleton, L. M. (2013). Using multilevel structural equation modeling techniques with complex sample data. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 521–562). Charlotte, NC: Information Age.

Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40,* 41–54. http://dx.doi.org/10.1016/j.cedpsych.2014.05.005

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797–811. http://dx.doi.org/10.1037/0022-3514.69.5.797

Sternberg, R. J., & Grigorenko, E. L. (Eds.). (2003). *The psychology of abilities, competencies, and expertise*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511615801

Stokes, D. E. (2011). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution Press.

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The new era of mixed methods. *Journal of Mixed Methods Research, 1,* 3–7. http://dx.doi.org/10.1177/2345678906293042

Thorndike, E. L. (1910). The contribution of psychology to education. *Journal of Educational Psychology, 1,* 5–12. http://dx.doi.org/10.1037/h0070113

van Deursen, A. J., & van Dijk, J. A. (2014). The digital divide shifts to differences in usage. *New Media & Society, 16,* 507–526. http://dx.doi.org/10.1177/1461444813487959

VanSledright, B., & Maggioni, L. (2016). Epistemic cognition in history. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 128–146). New York, NY: Routledge.

Walker, H. M. (1958). The contributions of Karl Pearson. *Journal of the American Statistical Association, 53,* 11–22. http://dx.doi.org/10.1080/01621459.1958.10501419

Weinstein, A., & Lejoyeux, M. (2010). Internet addiction or excessive internet use. *The American Journal of Drug and Alcohol Abuse, 36,* 277–283. http://dx.doi.org/10.3109/00952990.2010.491880

Wentzel, K. R., Barry, C. M., & Caldwell, K. A. (2004). Friendships in middle school: Influences on motivation and school adjustment. *Journal of Educational Psychology, 96,* 195–203. http://dx.doi.org/10.1037/0022-0663.96.2.195

Westera, W. (2011). On the changing nature of learning context: Anticipating the virtual extensions of the world. *Journal of Educational Technology & Society, 14,* 201–212.

Whitehead, A. N. (1967). *The aims of education*. New York, NY: Mentor. (Original work published 1929)

Wineburg, S. S., Martin, D., & Monte-Sano, C. (2012). *Reading like a historian: Teaching literacy in middle and high school history classrooms*. New York, NY: Teachers College Press.

Winne, P. H., & Baker, R. S. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining, 5,* 1–8.

Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review, 20,* 429–444. http://dx.doi.org/10.1007/s10648-008-9080-9

Received October 24, 2016
Revision received February 10, 2017
Accepted February 13, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 163–173

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000267

Educational Psychology: A Future Retrospective

Karen R. Harris
Arizona State University and Learning Sciences Institute Australia, Australian Catholic University

In my response to Alexander's (2018) paper marking the 125th anniversary of the American Psychological Association and the field of educational psychology, I have taken the perspective of a member of our discipline from some time in the future who is contributing to a larger work looking back at the history and development of our field (thus, a "future retrospective"). As this "future author," I focus on Alexander's (2018) article and selected developments in our field and more broadly since 2018. Two of the five thematic areas of influence that had established an enduring legacy for the field identified by Alexander are the primary focus: (a) interdisciplinary and cross-disciplinary inquiry and (b) evidence-based practice (EBP). The concepts of theoretical integration and theoretical integrationists are discussed in relation to these themes. Early barriers to interdisciplinary approaches, including paradigm wars and a proliferation of false dichotomies, are noted. The emergence of complexity sciences and a complex systems framework for understanding learning and development is discussed, leading to deeper understanding of the unique social and historical context that shaped and informed our work in the second decade of the 21st-century, as well as the multifaceted context we work within today. Given the interwoven nature of the five thematic areas identified by Alexander, however, aspects of the other thematic areas and Alexander's thoughts on the future of educational psychology are also encountered. I concur with Alexander in hoping that her paper and the responses to it generate discussion.

Educational Impact and Implications Statement

In this article, I take the perspective of an educational psychologist from some time in the future who is contributing to a larger work looking back at the history and development of our field. The focus, in part, is on the future of interdisciplinary and cross-disciplinary inquiry and evidence-based practice and the contribution of educational psychologists and others to this future. In addition, how current issues and developments, including paradigm and social justice wars, an argument culture, and the emergence of complexity sciences and a complex-systems framework for understanding learning and development impact the future of education are explored.

Keywords: learning, interdisciplinary, evidence-based practice, complexity science, paradigms

As noted in the abstract, what follows is a "future retrospective" regarding the field of educational psychology. I have taken the perspective of a member of our discipline from some time in the future who is contributing to a larger work looking back at the history and development of our field. As this "future author," I focus on Alexander's (2018) article and selected developments in our field and more broadly since 2018.

Karen R. Harris, Division of Educational Leadership and Innovation, Mary Lou Fulton Teachers College, Arizona State University, and Research Professorial Fellow, Learning Sciences Institute Australia, Australian Catholic University.

Appreciation is expressed to Ida Malian for noting the importance of considering the difference between "the future educational psychology" and "the future of educational psychology," and to Debra McKeown and Ralph P. Ferretti for their feedback on an earlier version of this article.

Correspondence concerning this article should be addressed to Karen R. Harris, Division of Educational Leadership and Innovation, Mary Lou Fulton Teachers College, Arizona State University, P.O. Box 871811, Tempe, AZ 85287-4501. E-mail: karen.r.harris@asu.edu

For my contribution to this retrospective on the field of educational psychology, I have been asked to look back at the history and development of the field, with particular reference to an article published by Alexander in 2018. As many readers know, in her article marking the 125th anniversary of the American Psychological Association and the field of educational psychology, Alexander identified five thematic areas of influence that had established an enduring legacy for the field at that time. She then projected potential paths the field might take in the next 25 years related to these themes.

Society and the field of educational psychology have changed a great deal since the second decade of the 21st century (20teens) when Alexander (2018) wrote that article. Nonetheless, the five areas she identified remain not only relevant but critical to our field: (a) the "psychologizing" of education (bringing education into the field of science), (b) interdisciplinary and cross-disciplinary inquiry, (c) learning as a theoretical and empirical core, (d) individual differences and their measurement, and (e) evidence-based practice (EBP). In the space allotted here, I focus primarily on two of these five themes in terms of our field today: interdisciplinary and cross-disciplinary inquiry and EBP, explor-

ing how developments since the late 20teens and 2020s have impacted our field today. Given the interwoven nature of the five themes, as noted by Alexander, some aspects of the other themes will also be encountered.

Contextual Considerations

Thinking back a bit further than 2018, in 1920 the population of the United States was 105,710,620, and farmers constituted 27% of the labor force. By 1990, the United States population was 246,081,000 and farmers made up 2.6% of the labor force (Economic Research Service, 2000). Our field and our society today look quite different from those from the 20teens, as indeed the United States and the field of education of the 1990s would have looked to those of the 1920s. When teaching the history of our field, I find that the documentary, *School: The American Public Education* (https://www.macfound.org/documentaryfilm/206/) offers insights from colonial times to the 21st century and may be of interest to readers who want to broaden this context and understanding of education before the 20teens further. Social and cultural changes since the 20teens need not all be summarized here, but critical developments since Alexander's (2018) article are briefly noted next. As Alexander stated, ". . . what will change dramatically are conditions within the world, in the educational experience, and in learners" (p. 154).

Development and Change

Our field today exists within a multifaceted larger context, as it did in 2018. As McCarty, Mancevice, Lemire, and O'Neil (2017) noted, "each generation of researchers works within a unique social and historical context that shapes and informs our work" (p. 7S). Social, economic, cultural, and related conditions in our country and the world have changed, with transactional relationships between many fields of research and these many forms of change, as well as other factors. It has not been a simple or straight path, but progress in health, education, economic mobility, and other aspects of social and economic development elegantly explained by Rosling (cf. www.ted.com/speakers/hans_rosling, 2006–2014) during Alexander's time has continued. In 2018, there were clear indicators that the complex "knot" of poverty could be untangled, the will to do so was growing, and some progress was being made (Cruz, Foster, Quillin, & Schellekens, 2015; Kena et al., 2016; Lamy, 2013). We have turned the tide on oppression of many forms (e.g., social, systematic, institutionalized, internalized) and of many types (e.g., class, disability, economic, gender, racial, religious, sexual), although there is more to be done. The course of our nation, and others, has come closer to living up to the creed of freedom, equality (e.g., economic, moral, political, social), justice, and humanity (cf. Allen, 2014; Glaude, Jr., 2016).

Today, for example, we look at the widespread existence of poverty and homelessness in the 20teens (even among children, with the number of homeless children attending public schools reaching 1.36 million by 2014; cf. Lamy, 2013; www.nn4youth.org/learn/how-many-homelss/) with the same incredulity and abhorrence with which we view the history and consequences of slavery (forms of which existed in our country and around the world beyond 2018). At the time of Alexander's (2018) article, the most recent Condition of Education report (Kena et al., 2016) indicated that 20% of United States students were living in poverty, with that number as high as 29% in Mississippi. These numbers, however, were seen by many as conservative and representing an outdated national poverty line (cf. Jiang, Granja, & Koball, 2017; *National Center for Children in Poverty: Measuring Poverty* http://www.nccp.org/topics/measuringpoverty.html).

Further, 24% of traditional public schools and 39% of charter schools were identified as high-poverty schools. Inequality in school funding, due in part to an outdated reliance on property prices, variation in per-pupil spending even within districts, and continued racial separation in schools, had not yet been powerfully addressed by 2018 (Ostrander, 2015; Spatig-Amerikaner, 2012). These high poverty schools often have fewer and poorer material resources and less-experienced teachers and leaders than their higher SES counterparts. Kindergartners entering these schools from socioeconomically disadvantaged households were significantly less ready for school, received lower test scores in the elementary grades, were more likely to drop out, and were less likely to enter higher education (Isaacs, 2012; Jiang et al., 2017).

In 2018, public schools enrolled 50 million students, 50% of whom were White and 9.3% of whom were students learning to speak English (Kena et al., 2016). Significantly higher percentages of Black, Hispanic, American Indian, and Alaska Native students attended high-poverty public schools than did White, Asian, and other students. Although 82% of public school students received a regular diploma and the dropout rate had reached 6.5%, this and previous Condition of Education reports made clear that the impact of educational disparities between White students and students of color was substantial. The effects of poverty, for children and adolescents of all races, on multiple forms of achievement well beyond the school years were indisputable by 2018 (cf. Dietrichson, Bog, Filges, & Jorgensen, 2017; Lamy, 2013; Nichols, 2016; Putnam, 2015; Wells, 2009).

Similarly, it is hard to understand how the situation and rights of the mentally ill were so widely disregarded in the 20teens and beyond. One in 17 individuals in the United States lived with a serious mental illness (e.g., schizophrenia, bipolar disorder, obsessive–compulsive disorder, substance-use disorder; prenatal exposure to drugs) in 2017. Further, 45% of homeless individuals were mentally ill, with 25% seriously mentally ill (https://mentalillnesspolicy.org/consequences/homeless-mentallyill.html). Although those with varying disabilities had experienced progress as a result of social rights movements such as special education and the American's with Disabilities Act of 1990 (https://www.dol.gov/general/topic/disability/ada), this group continued to battle oppression and stigmatization well past Alexander's (2018) time. As Glenn (2000) noted, however, the second civil rights movement and second-wave feminism in the second half of the 20th century, along with the successes of the disability-rights movement, created major legal, political, and social contextual changes that continue to provide a basis for ongoing change today.

Our societal battle against oppression by factors such as poverty, mental illness, disability, sexuality, and others is not complete; the challenges faced have been substantial (cf. Allen, 2014; Berliner & Glass, 2014; Glass, 2008; King, 2017; Lamy, 2013). Today, the impact of numerous and complex factors on social and economic progress over time continues to be studied and addressed. In the United States, for example, the economic stimulus-education spending that began in the 2000s (cf. Sparks, 2017) was one small

but important factor. Education spending today would have been difficult for those of the 20teens to foresee; complex models of economic progress have made clear the foundational role excellence in education for all plays in economic growth, and acceptance of this role is widespread. Further, educational psychology and the larger fields of education and psychology have been important players in progress toward what was termed social justice in the 20teens (for some insight to the hard-to-understand, nonsensical response of some against this movement for equal rights and opportunities at that time, see http://www.heritage.org/poverty-and-inequality/report/social-justice-not-what-you-think-it).

One final substantive contextual change since 2018 crucial here is the acknowledgment by the greater part of society that complex problems are the responsibility of the larger society rather than a few groups or institutions (cf. Brint, Turk-Bicakci, Proctor, & Murphy, 2009; Lamy, 2013; Lyall & Fletcher, 2013; Putnam, 2015; Wells, 2009). Before and beyond 2018, for example, teachers, schools, and colleges of education were frequently and sometimes fiercely blamed for complex societal outcomes beyond their control alone (cf. de Vink, 2015; Fuller, 2014; Lamy, 2013). As numerous researchers and reports concluded by 2018 (e.g., Berliner, 2009; Dietrichson et al., 2017; Goldstein, 2015; Putnam, 2015), factors outside of school exerted as much, or possibly more, influence on the life outcomes of children growing up in poverty as school factors. Addressing significant social problems, including oppression and poverty, requires that the majority of a society recognizes, advocates for, funds, and sustains meaningful improvements. We have reached this tipping point in the United States and other countries (with several countries there sooner than we were), but the balance yet totters and the movement toward a socially just society continues to need society's steadfastness and provision.

Who Qualifies as an Educational Psychologist?

Second, Alexander (2018) provided a provocative and thoughtful definition of what qualifies someone as a member of the educational psychology community given the increased focus on interdisciplinary and cross-disciplinary inquiry in the 20teens, and one often cited today. "What qualifies . . . individuals as members of the educational psychological community writ large is not that they hold a degree from an established educational psychology program, but that they share the mission of 'psychologizing' educational experiences," and share in the goal to "improve education, teaching, and learning (p. 149)." The diversity and number of fields today producing individuals who contribute to this mission were difficult to anticipate in 2018. By the late 2020s, in addition, other aspects of "becoming" an educational psychologist were also coming into play.

As educational psychology faculty worked more and more closely with faculty in school psychology and what were termed "general" and "special" education at that time (and with schools and communities, as Alexander called for), recruitment into educational psychology and related fields underwent significant changes. Skilled teachers and educational leaders became a key target for recruitment into our and others' doctoral programs. As life and careers continued to become longer, going straight through the undergraduate/masters and doctoral degrees became somewhat less common in our field, and experience as an effective teacher or educational professional became more sought after in doctoral candidates.

New Roles for Educational Psychologists

The third factor considered here is related to the second one, and to Alexander's (2018) well-stated concern regarding the burgeoning of terminology that plagued educational research at that time, and for some time to come. The potential for creating "incommensurable ways of speaking" (Vealey & Rivers, 2014, p. 174) was recognized across disciplines from multiple fields and specializations. As the complexity of the field grew, so did the types and forms of specializations within educational psychology. Whereas Alexander "undertook the role of historian" (p. 147) in identifying the five thematic areas noted earlier, what the formal roles of "historians" of educational psychology encompass today would have been difficult to envision in 2018. For example, one important role has been to track, analyze, synthesize, differentiate, and communicate bodies of terminology and their relationships to aspects of research and practice across and within fields. Distinguished careers have been committed to this work, and the impact on the field has been pivotal.

Further, historians have created reviews and integrations of theoretical, methodological, epistemological, and other perspectives in our field, helping educational psychologists gain insight into what the members of the field have learned and accomplished from multiple perspectives, and stimulating further work. This work has also meaningfully reduced the "re-creation of the wheel" that was clearly seen by Alexander's time, for example, reducing the introduction of new terms for already well-recognized constructs. Although new and refined constructs important to the field are welcome, researchers today must take care to collect the data and provide evidence regarding what is new and why it is important. The work of historians has been foundational here, as in the past it was possible for researchers to lack deep knowledge of work outside of their areas.

Other career paths for educational psychologists would have been hard to foresee in the 2020s, but are central to our field today. For example, although directors of research and research offices existed in school systems in the 20teens and earlier, that role has expanded today to encompass accomplished teams of researchers working within school systems and state and national education offices. These members of our field publish high-quality work and compete for funding. Educational psychologists have become critical members of policy-making groups and politician's teams, working to interpret research and research syntheses and promote their impact; develop deeper understanding of continuing research needs; and assist in interpreting and translating research for practitioners, families, organizations, and others (cf. Garcia, 2017; Tseng, 2012).

Having established aspects of context, the two relevant thematic areas on which I chose to focus are addressed next. Although I address each separately, these two areas are not independent of each other, and illustrate the interwoven nature of the five thematic areas identified by Alexander in 2018.

Challenges to an Interdisciplinary Beginning

Alexander (2018) described educational psychology as, by design or necessity, interdisciplinary from its inception. Belief in the

value of scientific inquiry applied to education, or what she termed the psychologizing of education, brought together scholars from diverse disciplines at the inception of the field, including philosophy, psychology, medicine, and more. From the beginning of the field, recognition of the complex challenges of education made it clear that multiple perspectives were needed.

Paradigm Wars

However, for many decades, there were limits to the extent to which such a conception of the field and interdisciplinary work were expressed. How can there be interdisciplinary work without disciplines, and without strong development of knowledge bases within these disciplines? In the first century of our field, a number of theoretical, epistemological, and methodological approaches to the understanding of learning, identified by Alexander (2018) as the core construct of our field, emerged and were intensely studied, and much was learned. These differing areas of study, however, sometimes vied for primacy or were monistic, which was true not only in education, educational psychology, school psychology, and psychology, but in many other scientific fields as well (cf. Hall, Yip, & Zarate, 2016a, 2016b; Harris, 1990; Harris & Graham, 1994; Miller et al., 2008; Schwartz, Lilienfeld, Meca, & Sauvigne, 2016; Pellmar & Eisenbery, 2000; Staats, 2005, 2016).

Historians have noted that, in the late 1960s, movement toward increased ideological and technological integration was becoming evident across a number of fields, including ours, as a result of forces in our culture and society (cf. Becher & Trowler, 2001; Boyer, 1990; Brint et al., 2009; Jacobs, 2014; Moran, 2010; Slavicek, 2012). By 1989, however, Gage wrote his seminal piece on the paradigm wars and their effect on the study of learning. Concerns for the future of interdisciplinary approaches were evident, as identification of fragmentation in psychology, education, and educational psychology grew. As Bruner (1990) noted, "Too often they [the parts of psychology] seal themselves within their own rhetoric and within their own parish of authorities" (p. ix), making it difficult to communicate with others "dedicated to the understanding of mind and the human condition" (p. x). As members of our field know, these paradigm wars intensified for several decades, leading some to talk about epistemological violence (Hall et al., 2016b) and paradigm as "paradogma" (Kirschner, 2014). Methodological wars were accompanied by wars over theories; there were also reading, math, writing, and science learning wars where theoretical, methodological, and epistemological views continued to repeat and collide (cf. Harris, 1990; Harris & Graham, 2016; Miller et al., 2008).

Social Justice Wars

At the time of Alexander's article in 2018, the social justice wars among some in diverse fields were also reaching a peak, such that members of some theoretical/research/methodological groups accused other groups of not being concerned with social justice, or worse. Proponents of some theories argued that "their" theory was the only theory that recognized the social justice issues and the only theory from which social justice could be meaningfully addressed. Examples of fragmentation based on social justice wars by 2018 ranged as follows.

- A conference session that disintegrated into a yelling match between scholars accusing each other of racism/lack of truly caring for children because one used the term disadvantaged (initially introduced to soften the stigma of poverty) while the other used the term at-risk (a term introduced by some to address perceived stigma in the word disadvantaged and later also seen as stigmatizing).
- The defamation of researchers who had worked in schools for decades and developed EBPs resulting in meaningful improvements in learning for many students (including students learning English and students in high-poverty schools) as racists, reasoning that their approach to teaching and learning was based on theories that "necessitated a rejection of who these children are and their culture."
- The pronouncement that only research based on a particular paradigm would be accepted in certain journals, including a prominent science-education research journal (cf. Harris & Graham, 2016, 2017; Jacobson, Kapur, & Reimann, 2016; King, 2017; Kirschner, 2014).

Members of disciplines once hailed as "dappled" (a discipline composed of manifold disciplinary, theoretical, and onto-epistemological perspectives; Lauer, 1984) moved to virulently contesting targeted theories or proclaiming dominant theories (cf. Bazerman, 2008; Prior, 2006; Vealey & Rivers, 2014). It is difficult now to understand how some researchers dedicated to understanding the human condition, learning, and development could see others, in fact their natural allies (cf. LaBoskey, 1998; King, 2017), as not only uncommitted to children, families, teachers, schools, communities, and social justice, but as enemies. Sadly, and ironically, these scholars who saw their preferred theory(s) as the only foundation for improving social justice expended a great deal of energy building straw men and perpetuating false dichotomies and prejudices. In contrast, many researchers, parents, teachers, and others strongly believed that improving teaching and learning for students in marginalized or impoverished situations was one powerful, albeit not sufficient, avenue toward social justice (cf. Dietrichson et al., 2017).

False Dichotomies and an Argument Culture

A marked and often noted aspect of Alexander's time was theoretical, paradigmatical, and political polarization across areas of education resulting in the proliferation of false dichotomies (cf. Garcia, 2017; Harris, 1982; Harris & Graham, 1994, 2016; King, 2017; Jacobson et al., 2016; Resnick, 1987). Recognition of the nature and impact of these false dichotomies (or in some cases, trichotomies or larger) and accompanying disparaging rhetoric was evident in the 20teens and much earlier, yet they persisted for some years to come. Misleading and false dichotomies, polarized positions, and their negative outcomes were prevalent beyond Alexander's time.

The argument culture. Tannen (1998), a prominent sociolinguist whose work focused on observing and explaining language and its role in human relations during Alexander's time, wrote a thoughtful and provocative book based on several years of study exploring "the argument culture" and "war of words" that was becoming pervasive in the United States. Her work indicated that many in the United States had adopted perspectives leading to an argument culture, a culture in which it had become common to approach the world in an adversarial frame of mind. Public dialogue had frequently become approached as a fight to be won,

rather than an opportunity for discourse. The quest to prove that one is right rather than explore and understand other viewpoints, including potential weaknesses in others' positions, had a number of negative consequences, including polarized views and false dichotomies. As Tannen noted, "Public discourse requires *making* an argument for a point of view, not *having* an argument—as in having a fight" (p. 4).

Tannen (1998), however, also stated that "Sometimes passionate opposition, strong verbal attack, are appropriate and called for" (p. 7). Throughout her book, she provided illustrations of when it is important and necessary to argue "for right against wrong or against offensive and dangerous ideas and actions" (p. 8). What then, did her research and analysis indicate that differentiated between reasoned, passionate opposition and an argument culture? Tannen's study of language and its impact on relations and communication encompassed fields including the press, politics, litigation, gender, public education, academia, and culture. Across these foci she identified critical aspects of the argument culture, including the use of criticism, attack, and opposition as the primary, or only, means of responding to people or ideas. The goal was to win, rather than to listen, understand, and learn. Rather than thinking critically and responding to viewpoints other than one's own, which requires analysis and interpretation, the default position was criticism and the belief that for one to be right, others must be wrong. The argument culture can lead individuals or groups to distort facts and others' positions, seize on irrelevant details, deny facts that support an "opponent's" views, oversimplify issues and viewpoints, limit thinking and knowledge rather than expand it, obscure aspects of divergent work or viewpoints that in fact overlap and have the potential to enlighten understanding, and mount unfair, even vicious attacks on professionals that take time away from meaningful work and create harm. Many scholars in multiple fields rejected this behavior and sought productive relationships and multiple views, making significant progress in multiple fields. Tannen, however, summarized continuing concerns by many in the 2020s.

> Of course it is the responsibility of intellectuals to explore potential weaknesses in others' arguments . . . But when opposition becomes the overwhelming avenue of inquiry—a formula that *requires* another side to be found or a criticism to be voiced; when the lust for opposition privileges extreme views and obscures complexity; when our eagerness to find weaknesses blinds us to strengths; when the atmosphere of animosity precludes respect and poisons our relations with one another; then the argument culture is doing more damage than good. (p. 25)

False dichotomies. As an academic, Tannen (1998) noted that, too often, the culture of the academy encouraged individuals or groups to position their work in opposition to others' work and then set out to prove others wrong, precluding or obstructing deep inquiry and meaningful insights in the ways previously described. As a result, she noted, "Straw men spring up like scarecrows in a cornfield" (p. 269). A sampling of polarizing, simplistic, and misleading false dichotomies in the field of education evident at the time of Alexander's (2018) article is included in Table 1. In 1998, LaBoskey voiced the concern raised by the prevalence of such an approach to the field of education well.

Table 1
*Illustrative False Dichotomies, Trichotomies, (or More) Evident by 2018**

Constructs
Instruction/Instructivist vs. Construction/Constructivist
Teacher Centered vs. Learner Centered
Empiricism vs. Holism
Interdisciplinary vs. Transdisciplinary
Emergent Learning vs. Single Trial Learning
Educational Psychology vs. Learning Sciences
Cognitive vs. Situative
Deficit Model vs. Whole Child Model
Cultural Similarities vs. Cultural Uniqueness
Discourse vs. Dialogue vs. Teacher Talk
Social vs. Ecological vs. Sociocultural
Evolutionary Processes vs. Neurophysiological Processes vs. Situated Processes vs. Sociocultural Processes

* Selected references: Blank, 2002; Box, Skoog, and Dabbs, 2015; Dvorakova, 2016; Hall, Yip, and Zarate, 2016a, 2016b; Harris and Graham, 1994; Harris and Pressley, 1991; Jacobson, Kapur, and Reimann, 2016; Schwartz, Lilienfeld, Meca, and Sauvigné, 2016; Staats, 2005, 2016.

> [Dichotomies] characterizes with chilling accuracy our common approach to educational problem-solving. It is such dichotomous thinking . . . that contributes to the continuation of our difficulties . . . We spend much too much of our limited time, energy, and resources debating these "false dichotomies." Instead of coming together as a community deeply concerned about the future of our children, we make artificial enemies of one another. (p. 39)

Selected illustrations. Harris (2014; Harris & Graham, 2016, 2018) identified one long-enduring false dichotomy that negatively impacted our field as the forced choice between instructed and constructed knowledge. She described the belief that learning to read and write should parallel how we learn to talk—and that learning to talk is a "natural process" rather than an instructed, scaffolded process. By 2018, this viewpoint had spawned widespread approaches to teaching reading, writing, math, and science based on the belief, not validated by research, that learning would develop naturally through rich immersion in authentic learning environments, with little to no explicit instruction. For some, "teaching" was seen as a "dirty word" (Harris & Graham, 1994). This approach was not highly successful for a large number of students over several decades (cf. Harris & Graham, 2016; National Reading Panel, 1999). Harris (2014 and Harris & Graham, 2016, 2018) argued that although the importance of a rich environment and active involvement of the child in language development was indisputable, those who had raised or loved a baby also observed how all of those who interact with that baby contribute to scaffolding language development. Total strangers will make noises for babies to hear and imitate; parents and friends prompt babies to produce sounds and words and then assist them in pronunciation; siblings and others show or explain what a word means; and so on. Adults, siblings, and peers interact explicitly with babies and young children in myriad ways, from birth through childhood, to help support learning to speak and language development. As she concluded, there is perhaps no more explicitly scaffolded and supported learning experience for most of us in our lives.

As research made clear well before 2018 and into the 2020s, inclusion of supported, explicit aspects of instruction is not incom-

patible with constructivist and other views of learning that emphasize active learning and construction of knowledge (Harris, 1982; Harris & Pressley, 1991). Knowledge transformation and construction occur across approaches to teaching and learning, making a forced choice between constructed and instructed learning unnecessary (cf. Resnick, 1987). Learning to read and to write for most children, and other forms of complex learning, requires immersion, dialogue, meaningful activity, collaboration, explicit scaffolding, instruction, practice, and feedback. In addition, complex learning requires teachers and researchers' careful consideration of many more factors, such as affective, behavioral, cognitive, contextual, cultural, developmental, and social factors (cf. Harris & Graham, 2009, 2017; Miller et al., 2008; Staats, 2005).

Polarizing views and false dichotomies reduce complexities, which evolve over time and contexts, to simplistic labels. For example, the reference to individuals (whom they often did not know) who worked from targeted paradigms or in particular fields, or entire fields (such as special education), as working from a "deficit model" obscured the complex contextual issues (e.g., political, social, economic, legal, definitional) those with and without disabilities faced as they fought this civil rights battle over multiple decades (Harpur, 2012; Kauffman, 2009; Pelka, 2012; Winzer, 2009).

False dichotomies, however, do not preclude extreme viewpoints from existing. Indeed, there were those in the 2020s and beyond who saw individuals with what was termed a disability as defined by their challenges rather than their strengths, who could not understand the difficulties in valid definition or identification of disabilities, who treated individuals with disabilities as passive recipients of intervention, and who failed to see the child or adult as a unique individual. This perspective historically resulted in shunning and isolating those with meaningful differences and challenges from society.

False dichotomies, rather, are "false" because they present oversimplified extremes and obfuscate complexity, although they may result from passionate and well-intended principles. The history of special education makes it clear that those striving for understanding of disabilities and individuals with disabilities were working from the perspective of "the whole child/adult" (cf. Harpur, 2012; Kauffman, 2009; Pelka, 2012; Winzer, 2009). Those involved in the disability-rights movement and special education were committed to understanding individuals' strengths, unique attributes, and contributions while also developing means of working with communities, families, children, and adults to address challenges and needs. The path to reform is seldom straight, simple, or foreseeable; mistakes, unforeseen consequences, and new developments occur and must be resolved. When solutions are attempted for complex problems, that effort can impact the understanding of the problem and may reveal or create additional problems (Jones, 2011), as can be seen in the history of special education. As Tannen (1998) and LaBoskey (1998) noted, engaging in oversimplified false dichotomies is an impediment to the understanding and progress needed to address complex problems.

The limitations, challenges, and barriers encountered in the fight for the rights of many, including those with disabilities, have taken society far longer to address than hoped. An important component of the progress we have made is the recognition of, and intolerance for, false dichotomies and an argument culture. Scholarship demands respect and careful consideration of multiple viewpoints, deep inquiry, robust understandings, and reflective action.

Recurrence, Persistence, and Acceptance of Interdisciplinary Approaches

Although it would be spurious to claim that paradigm wars, false dichotomies, and related issues no longer exist in our field, here again we have come a long way. And although the call for an interdisciplinary focus is threaded throughout the history of not only educational psychology, but many other disciplines, it was not until the 2020s that integration of diverse perspectives in the movement to improve learning became a major force in our field. Alexander (2018) noted that "interdisciplinary and cross-disciplinary inquiry seemingly arises organically when the nature or complexity of the problems to be addressed demand it" (p. 149). Others from multiple disciplines shared that view, both during Alexander's time and much earlier (cf. Boyer, 1990; Brint et al., 2009; Dubin, 1978; Jacob, 2015; Jacobs, 2014; Jacobson et al., 2016; McCarty et al., 2017; Lyall & Fletcher, 2013; Miller et al., 2008; Morris & Reardon, 2017; National Academy of Science, 2005; Pellmar & Eisenbery, 2000; Slavicek, 2012).

Although a plethora of terms for integrative approaches existed by 2018, including cross-disciplinary, multidisciplinary, critical interdisciplinarity, eclectic interdisciplinarity, transdisciplinary, and others (Jacob, 2015; Jacobs, 2014; Lyall & Fletcher, 2013; National Academy of Science, 2005; Pellmar & Eisenbery, 2000), I use the term interdisciplinary here to refer to the overarching concept of individuals from two or more areas of expertise working together to address complex, significant current and emerging challenges. Expertise across domains, including methodologies and epistemologies, allows a more powerful understanding of such challenges and problems, thus more powerful responses. No single domain, methodology, or epistemology is empowered over others in such an approach. Rather, this approach allows us to leverage problem-based, collaborative teams that draw from theories, designs, and methods appropriate to the complex issues and challenges (Jacob, 2015; Jacobs, 2014; Lyall & Fletcher, 2013).

Closely related to the emergence of interdisciplinary approaches to complex problems in the 20teens were discussions of the power, and potential weaknesses, of the union of knowledge among a community, referred to by terms such as collective intelligence, group mind, collective problem solving, collective impact, collective cognition, and distributed knowledge (Allen, 2014; Fagin, Halpern, Moses, & Vardi, 1995; Hung, 2013; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). Some of the many factors related to the current level of interdisciplinary work in our field and others, are briefly noted here.

In her history of interdisciplinarity across the social sciences, natural sciences, humanities, and professions, Klein (1990) argued that a complex network of historical, social, psychological, political, economic, philosophical, and intellectual factors are embodied in interdisciplinary work. It might be added, further, that it takes a critical mass of these factors coming together in the first place to set the stage for interdisciplinary approaches to take root and thrive (cf. Jacob, 2015; Lyall & Fletcher, 2013). For example, this approach to research will often require substantial funding and time (cf. Lyall & Fletcher, 2013; National Academy of Science, 2005; Pellmar & Eisenbery, 2000). By the time Alexander's

(2018) article was published, calls for and sources of funding for interdisciplinary research, and the development of interdisciplinary researchers, to address multifaceted social problems were beginning to multiply across governmental, private, and other organizations in numerous countries, including the United States (e.g., https://ies.ed.gov/funding/ncer_rfas/predoctoral.asp; https://www.nsf.gov/od/oia/additional_resources/interdisciplinary_research/; http://www.nordp.org/funding-opportunities; http://www.rcuk.ac.uk/funding/gcrf/interdisciplinary-research-hubs-to-address-intractable-challenges/; https://www.britac.ac.uk/sites/default/files/Crossing%20Paths%20-%20Full%20Report.pdf; http://www.rwjf.org/en/library/funding-opportunities/2017/interdisciplinary-research-leaders.html).

At the same time, as noted previously, awareness and rejection of factors that had inhibited interdisciplinarity previously, such as paradigm wars and false dichotomies, were taking root across graduate programs in all areas of education and numerous related fields. Graduate programs across diverse fields created and refined degree programs that prepared graduates to address social, economic, and political issues from interdisciplinary viewpoints, while at the same time focusing on development of deep disciplinary knowledge and expertise (Brint et al., 2009; Jacob, 2015; Jacobs, 2014; Schmidt et al., 2012).

The emergence of generations of scholars philosophically, psychologically, and intellectually prepared to work collaboratively to address varying challenges situated in local, international, and/or interorganizational frameworks was fundamental to the growth of interdisciplinarity. Such undertakings, however, took time, and the pace of interdisciplinary progress was sometimes frustrating in education, as it was in other fields. However, as critical masses of scholars emerged and broader political, attitudinal, and economic changes in society occurred (some of which were noted previously), meaningful problem solving produced undeniable progress, leading to further investments.

EBP

As discussed earlier, new roles for educational psychologists have developed and thrived since 2018, including roles in schools and policy arenas. As the transformative power of EBPs in medicine, agriculture, technology and other fields became clear, this movement also gained critical momentum in education and educational psychology and was a natural locus for interdisciplinary work (cf. Cook, Smith, & Tankersley, 2012; Garcia, 2017; King, 2017; Sparks, 2017). For far too many decades, the adoption of curriculum and teaching methods in our country and others (too often following our lead) had been driven by "Pied Pipers"[1](Harris & Graham, 2016). Engaging and effective speakers and writers with an instructional method or approach to sell to their colleagues, communities, families, schools, and parents, but no evidence to back their claims regarding that actual approach or product, were far too common (cf. https://www2.ed.gov/rschstat/research/pubs/rigorousevid/rigorousevid.pdf). As early as 1972, Bloom noted the following.

> In education, we continue to be seduced by the equivalent of snake-oil remedies, fake cancer cures, perpetual-motion contraptions, and old wives' tales. Myth and reality are not clearly differentiated, and we frequently prefer the former to the latter. We have been innocents in education because we have not put our house in order . . . We need to be much clearer about what we do and do not know so that we do not continually confuse the two. If I could have one wish for education in the next decade it would be the systematic ordering of our basic knowledge in such a way that what is known and true can be acted on, although what is superstition, fad, and myth can be recognized as such and used when there is nothing else to support us in our frustration and despair. (pp. 333–334)

Much as in the field of medicine, the EBP movement faced many challenges and setbacks, yet by the late 20teens, was beginning to thrive (cf. Cook et al., 2012; Pellmar & Eisenbery, 2000). As Tseng (2012) reported, a framework for policymakers and practitioners to understand the uses of educational research, develop the knowledge and ability to evaluate the implications of research, understand the many limitations of single studies, and continue to identify resources to assist them as necessary when evaluating research was developing.

> The framework describes the ways policymakers and practitioners define, acquire, interpret, and ultimately use research. Relationships are vital conduits for acquiring research. When confronted with questions about a program or reform, agencies and legislators often turn to trusted peers and intermediaries. Translation is also key. Because research does not speak for itself, policymakers and practitioners must always interpret its meaning and implications for their particular problems and circumstances. This means that identifying the right translators and creating productive conditions for translation are critical. (p. 1)

Educational psychologists have become indispensable members of teams continuing work in translation and impact. They have also been key players in identifying areas of "user-inspired research" based on engagement with communities, families, schools, teachers, students, and others (Bulterman-Bos, 2008; Stokes, 1997), while contributing to overall development of theory and knowledge. Effective partnerships between researchers, policymakers, and others invested in education, families, communities, our students, and our schools were critical in establishing EBPs as a transformative force in education.

Theoretical Pluralism and Theoretical Integrationists

As noted, concern for fragmentation in education and psychology was clear in 2018 and had been for many years (Boyer, 1990; Dubin, 1978; Jacobson et al., 2016; Staats, 2005). Interdisciplinary inquiry and approaches recurred, persisted, and eventually became one bedrock of our field. Thoughtful, effective integration of diverse, validated approaches to learning, regardless of whether or not the disciplines from which they originated were viewed by some as discordant (such as affective, behavioral, cognitive, constructivist, sociocultural, and other approaches to teaching and learning), continues to be one key to the development of our field. Epistemological, theoretical, and methodological pluralism (cf. Jacobson et al., 2016; Miller et al., 2008; Staats, 2005) became vital forces in the development of our field. As Dubin (1978) argued in his book, *Theory Building*, contiguous problem solving allows interdisciplinary efforts, based on disciplinary research, to

[1] A short story based on a legend from the town of Hamelin, Germany, in 1284. A piper dressed in multicolored ("pied") clothing is angered by the town's citizens and uses his instrument's magical power to lure the town's children away, never to be seen again.

add up in a way not otherwise likely. When we treat competing viewpoints with thoughtfulness and respect, a powerful repertoire for teaching and learning across the life span develops. This does not negate the importance of competition in advancing our thinking and research contributions (Dubin, 1978; Hall et al., 2016a, 2016b; Harris, 1982; Harris & Graham, 2017). Thus, we continue our commitment to disciplinary research, including basic research, as another bedrock of our field.

As many had noted decades earlier and in the 20teens, there are no panaceas in teaching and learning, or in research on teaching and learning. Single theories, including those prominent in Alexander's (2018) time, simply could not capture the complex nature of learning and the diversity among learners. Nor could any single theory address all of the challenges faced by learners, their teachers, and their families, schools, and communities. As is evident today, good instruction does not require a forced choice between competing theories, but rather a triangulation across and integration of the evidence from various theories, perspectives, and lines of research. Learning is a complex process that relies on development across diverse learners in multiple areas. Further, far before the time of Alexander's (2018) article, it was clear that all major theories of learning in our field embraced meaningful learning in educationally purposeful, open, just, disciplined, caring, and celebrative communities. It was also clear that skillful and enthusiastic teaching is critical, and that critical attributes of effective teachers and characteristics of effective instruction (cf. Brophy, 1979; Good & Brophy, 1997) belong to no single theory, but rather are supported by many (Harris, 1982, 2014; Harris & Graham, 2018).

Theoretical Integration and Theoretical Integrationists

The movement in teaching and learning toward theoretical integration in research and practice relied in part on purposeful triangulation across theories, identifying critical constructs (although named differently) that shared much in common. In many cases, deeper understanding of such critical constructs is augmented by multiple theoretical perspectives. Such critical constructs in teaching and learning today include developing deep and broad understanding of learners; what is to be learned; how this understanding of the learner and what is to be learned work together (arguably evident across multiple theories and approaches to teaching and learning by Alexander's time); how others play a role in learning; how school factors interact with learning; how family, culture, environment, and community influence learning; learning in and out of school; the role of learning in development; and more.

Although robust lines of research (from basic to applied) continue addressing established and emerging theories today, EBPs based on theoretical integration have proven instrumental to powerful teaching and learning. By the time of Alexander's (2018) article, the concept of theoretical integrationists was gaining ground. Theoretical integrationists (cf. Harris, 2014, 2016)

- Believe in all children and their futures
- Reject false dichotomies, prejudice, and straw men; treat diverse participating theories with thoughtfulness and respect
- Believe interdisciplinary relationships, built on trust and respect, are essential to the future of educational psychology, students, and society
- Focus on how knowledge is constructed and instructed; and on cultural, social, school, classroom, family, and community factors that impact learning and development, and
- Believe understanding and integrating what we know across theories, methods, epistemologies, and paradigms will allow us to advance the field by assisting policymakers and practitioners to define, acquire, interpret, adapt, and ultimately bring proven practices to scale.

Complexity Science and Complex Systems Approaches

At the same time as interdisciplinarity took root in research and the development of researchers, and as the EBP movement and theoretical integration took hold and began to thrive, another new approach to complex problems was evolving worldwide in both the physical and social sciences: complexity sciences (cf. Ackoff, 1974; Benham-Hutchins & Clancy, 2010; Jones, 2011). At the time of Alexander's (2018) paper, complexity theory and complexity sciences were emerging, yet many years would pass before they began to resemble the approaches to complex problems we use and continue to develop today.

Complexity science during Alexander's time, as today, was based on multiple theories and tools integrated across a range of disciplines (Jones, 2011). Complexity science referred to the study of complex systems and problems that are dynamic, often unpredictable, multidimensional, and include interconnected parts and relationships. Complex problems are often characterized by nonlinearity and are similar to the concept of a 'wicked problem' (Conklin, 2001) in that they lack a well-defined structure and straightforward solutions.

Ackoff (1974) stated that complex problems can be distinguished from simpler problems and puzzles, and referred to them as "messes." Complex problems, or messes, are characterized by multiple integrated and difficult to separate dimensions (e.g., cultural, economic, ethical, political, religious, technological). Differing, yet plausible and legitimate, perspectives and interpretations of the problem may also exist. Complexity science, and a related approach referred to as "theory of change" (Brest, 2010; Clark & Taplin, 2012; James, 2011; Vogel, 2012), were being explored by government agencies across multiple countries, international nongovernmental agencies, philanthropies, the United Nations, and other major organizations in response to national and international challenges such as health and health care, poverty, environmental issues, and organizational and political issues. The efficacious use of such approaches to change was only beginning to be understood, and it would be some time before their impact would be strongly felt in numerous fields, including education.

During Alexander's (2018) time, for example, growing attention was focused on a complex systems conceptual framework for understanding learning and development. This approach began producing new, innovative, and insightful ways of collecting and analyzing data that allowed the field to develop more sophisticated and complex models and theories of learning (cf. Bar-Yam, 2003; Byrk, 2015; Jacobson et al., 2016; Miller et al., 2008). Jacobson et al. (2016) stated the prospects of such approaches well:

> We hope principled theoretical considerations of learning as an emergent phenomenon in complex neural, cognitive, situative, social, and cultural systems will yield critically important insights of central relevance to our field that might not otherwise be possible with current perspectives and approaches. In addition, viewing the environments in which learning occurs as complex systems provides educational and learning researchers with powerful conceptual tools (e.g., computer modeling) that are being used by scientists in other areas of research. (p. 217)

Educational psychologists have contributed, and continue to contribute, theoretical and empirical components to this complex systems approach, not only in the areas above but across the broad areas of affect and behavior as well. Further, our field now plays a critical role in exploring individual learners as "complex systems" within larger complex systems.

Historians have identified a confluence of multiple aspects of development and change during Alexander's (2018) time, in both education and the larger social context, that contributed to progress and many of our accomplishments today, including interdisciplinary and integrative approaches; the emphasis on and growing use of EBPs for teaching and learning; the emergence of generations of scholars and educational leaders philosophically, psychologically, and intellectually prepared to work collaboratively to address complex challenges; and the slowly widening acknowledgment by more of society that complex problems, including poverty and issues of social justice, are the responsibility of the larger society rather than single groups or institutions such as teachers and schools. It would be some time before additional cultural and political factors, including the will to enact and bear the costs of many aspects of social and educational reform, joined in this trajectory. That time has come today, given the unique social and historical context that shapes and informs our work.

In 1990, Boyer identified "the probing mind of the researcher" (p. 18) as a vital asset to the academy and the world. Our "probing minds" continue to advance educational psychology and the larger field, allowing us to make progress in the "hardest science of all" (Berliner, 2002), one that "can demonstrate decade *x* treatment interactions, an occurrence almost unfathomable to most physical scientists" (Berliner, 2006, p. 20). To put our past, present, and future in perspective, consider all of the earth's history as if it occurred in just 24 hours. Single-celled algae appeared roughly 11 hr ago; multicellular organisms 7 hr ago. Aquatic animals arrived less than 4 hr ago, plants colonized land about 3 hr ago, and land animals began to appear just 2 hr ago. Dinosaurs showed up approximately 1 [1/2] hr ago and disappeared about 1 hr later. Earliest humankind appeared 2 min ago, and modern humans appeared only 1 s ago (http://jan.ucc.nau.edu/lrm22/lessons/timeline/24_hours.html). Although we often experience intense frustration about what we have not yet achieved, looking back at the history of our country, our field, and education around the world, with particular reference to Alexander (2018), provides context for recognizing progress made and impetus for continuing our work.

References

Ackoff, R. (1974). *Redesigning the future: A systems approach to societal problems*. New York, NY: Wiley.

Alexander, P. A. (2018). Past as prologue: Educational psychology's legacy and progeny. *Journal of Educational Psychology, 110,* 147–162. http://dx.doi.org/10.1037/edu0000200

Allen, D. S. (2014). *Our declaration: A reading of the declaration of independence in defense of equality*. New York, NY: Liveright.

Bar-Yam, Y. (2003). *Dynamics of complex systems*. New York, NY: Perseus.

Bazerman, C. (2008). *Handbook of research on writing: History, society, school, individual, text*. New York, NY: Taylor & Francis.

Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories* (2nd Ed.). Buckingham, England: Open University Press.

Benham-Hutchins, M., & Clancy, T. R. (2010). Social networks as embedded complex adaptive systems. *The Journal of Nursing Administration, 40,* 352–356. http://dx.doi.org/10.1097/NNA.0b013e3181ee42bc

Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher, 31,* 18–20. http://dx.doi.org/10.3102/0013189X031008018

Berliner, D. C. (2006). Educational psychology: Searching for essence throughout a century of influence. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 3–28). Washington, DC: APA.

Berliner, D. C. (2009, March). *Poverty and potential: Out-of-school factors and school success*. Boulder, CO and Tempe, AZ: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from http://nepc.colorado.edu/publication/poverty-and-potential

Berliner, D. C., Glass, G. V., & Associates. (2014). *50 myths and lies that threaten America's public schools: The real crisis in education*. New York, NY: Teachers College Press.

Blank, M. (2002). Classroom discourse: A key to literacy. In K. G. Butler & E. R. Silliman (Eds.), *Speaking, reading, and writing in children with language learning disabilities: New paradigms in research and practice* (pp. 151–173). Mahwah, NJ: Erlbaum.

Bloom, B. S. (1972). Innocence in education. *The School Review, 80,* 333–352. http://dx.doi.org/10.1086/443036

Boyer, E. L. (1990). *Scholarship reconsidered: Priorities of the professoriate*. Lawrenceville, NJ: Princeton Press.

Box, C., Skoog, G., & Dabbs, J. M. (2015). A case study of teacher personal practice assessment theories and complexities of implementing formative assessment. *American Educational Research Journal, 52,* 956–983. http://dx.doi.org/10.3102/0002831215587754

Brest, P. (2010). The power of theories of change. *Stanford Social Innovation Review, Spring,* 47–51.

Brint, S. G., Turk-Bicakci, L., Proctor, K., & Murphy, S. P. (2009). Expanding the social frame of knowledge: Interdisciplinary, degree-granting fields in American colleges and universities, 1975–2000. *The Review of Higher Education, 32,* 155–183. http://dx.doi.org/10.1353/rhe.0.0042

Brophy, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology, 71,* 733–750. http://dx.doi.org/10.1037/0022-0663.71.6.733

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Byrk, A. S. (2015). 2014 AERA distinguished lecture: Accelerating how we learn to improve. *Educational Researcher, 44,* 467–477. http://dx.doi.org/10.3102/0013189X15621543

Bulterman-Bos, J. A. (2008). Will a clinical approach make educational research more relevant to practice? *Educational Researcher, 37,* 412–420. http://dx.doi.org/10.3102/0013189X08325555

Clark, H., & Taplin, D. (2012). *Theory of change basics: A primer on theory of change*. New York, NY: Actknowledge. Retrieved from http://www.theoryofchange.org/wpcontent/uploads/toco_library/pdf/ToCBasics.pdf

Conklin, J. (2005). Wicked problems and social complexity. In *Dialogue mapping: Building shared understanding of wicked problems*. Hoboken, NJ: Wiley. Retrieved from http://cognexus.org/wpf/wickedproblems.pdf

Cook, B. G., Smith, G. J., & Tankersley, M. (2012). Evidence-based

practices in education. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Vol. 1. Theories, constructs, and critical issues* (pp. 495–527). Washington, DC: American Psychological Association.

Cruz, M., Foster, J., Quillin, B., & Schellekens, P. (2015). *World Bank group policy research note: Ending extreme poverty and sharing prosperity. Progress and policies.* Retrieved from http://pubdocs.worldbank.org/en/109701443800596288/PRN03Oct2015TwinGoals.pdf

De Vink, C. (2015, June 16). Teachers have become the popular scapegoats for society's ills. *Chicago Tribune*. Retrieved from http://www.chicagotribune.com/news/opinion/commentary/ct-blame-teachers-helen-keller-anne-sullivan-perspec-0617-20150616-story.html

Dietrichson, J., Bog, M., Filges, T., & Jorgensen, A. K. (2017). Academic interventions for elementary and middle school students with low socioeconomic status: A systematic review and meta-anlaysis. *Review of Educational Research, 87,* 243–282. http://dx.doi.org/10.3102/0034654316687036

Dubin, R. (1978). *Theory building*. New York, NY: Macmillan.

Dvorakova, A. (2016). The cultural psychology endeavor to make culture central to psychology: Comment on Hall et al. (2016). *American Psychologist, 71:*888–889, 2016. http://dx.doi.org/10.1037/amp0000053

Economic Research Service. (2000, September). *A history of American agriculture, 1607–2000* (ERS-POST-12.), Washington, DC: Author.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning about knowledge*. Cambridge, MA: MIT Press.

Fuller, E. J. (2014). Shaky methods, shaky motives: A critique of the National Council of Teacher Quality's review of teacher preparation programs. *Journal of Teacher Education, 65,* 63–77. http://dx.doi.org/10.1177/0022487113503872

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher, 18,* 4–10.

Garcia, D. R. (2017, January 11). The missing question in education. *Education Week, 36,* 20.

Glass, G. V. (2008). *Fertilizers, pills, & magnetic strips: The fate of public education in America*. Charlotte, NC: Information Age Publishing.

Glaude, E. S., Jr. (2016). *Democracy in black: How race still enslaves the American soul*. New York, NY: Crown Publishing.

Glenn, E. N. (2000). Creating a caring society. *Contemporary Sociology, 29,* 84–94. http://dx.doi.org/10.2307/2654934

Goldstein, D. (2015). *The teacher wars: A history of America's most embattled profession*. New York, NY: Anchor Books.

Good, T., & Brophy, J. (1997). *Looking in classrooms* (7th ed.). New York, NY: Longman.

Hall, G. C. N., Yip, T., & Zárate, M. A. (2016a). On becoming multicultural in a monocultural research world: A conceptual approach to studying ethnocultural diversity. *American Psychologist, 71,* 40–51. http://dx.doi.org/10.1037/a0039734

Hall, G. C. N., Yip, T., & Zárate, M. A. (2016b). Disciplinary perspectives on multicultural research: Reply to Dvorakova (2016) and Yakushko et al. (2016). *American Psychologist, 71,* 892–893, 2016. http://dx.doi.org/10.1037/amp0000124

Harpur, P. (2012). From disability to ability: Changing the phrasing of the debate. *Disability and Society, 27,* 325–337.

Harris, K. R. (1982). Cognitive-behavior modification: Application with exceptional students. *Focus on Exceptional Children, 15,* 1–16.

Harris, K. R. (1990). Paradigms. *Journal of Learning Disabilities, 23,* 586–587.

Harris, K. R. (2014). Message from the president. *NEP/15, Fall,* 1–2. Retrieved from http://apadiv15.org/wp-content/uploads/2014/12/NEP-Fall-2014.pdf

Harris, K. R., & Graham, S. (1994). Constructivism: Principles, paradigms, and integration. *The Journal of Special Education, 28,* 233–247. http://dx.doi.org/10.1177/002246699402800301

Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. *Teaching and Learning Writing, 1,* 113–135. http://dx.doi.org/10.1348/978185409X422542

Harris, K. R., & Graham, S. (2016). Self-regulated strategy development in writing: Policy implications of an evidence-based practice. *Policy Insights from the Behavioral and Brain Sciences, 3,* 77–84. Retrieved from http://dx.doi.org/10.1177/2372732215624216

Harris, K. R., & Graham, S. (2018). Self-regulated strategy development: Theoretical bases, critical instructional elements, and future research. In R. Fidalgo, K. R. Harris, & M. Braaksma (Eds.), *Studies in writing: Vol. X: Design principles for teaching effective writing: Theoretically and empirically grounded principles* (pp. 119–151). Leiden, the Netherlands: Brill Editions.

Harris, K. R., & Pressley, M. (1991). The nature of cognitive strategy instruction: Interactive strategy construction. *Exceptional Children, 57,* 392–405. http://dx.doi.org/10.1177/001440299105700503

Harris, K. R. (2016, August). *Self-regulated strategy development for writing: Confessions of an evidence-based practice*. Division 15 (Educational Psychology) Presidential address presented at the Conference of the American Psychological Association, Denver, CO.

Hung, W. (2013). Team-based complex problem solving: A collective cognition perspective. *Educational Technology Research and Development, 61,* 365–384. http://dx.doi.org/10.1007/s11423-013-9296-3

Isaacs, J. B. (2012). *Starting school at a disadvantage: The school readiness of poor children*. Washington, DC: Brookings Institution, Center on Children and Families. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/0319_school_disadvantage_isaacs.pdf

Jacob, W. J. (2015). Interdisciplinary trends in higher education. *Palgrave Communications, 1,* Article 15001. http://dx.doi.org/10.1057/palcomms.2015.1

Jacobs, J. A. (2014). *In defense of disciplines: Interdisciplinarity and specialization in the research university*. Chicago, IL: University of Chicago Press. http://dx.doi.org/10.7208/chicago/9780226069463.001.0001

Jacobson, M. J., Kapur, M., & Reimann, P. (2016). Conceptualizing debates in learning and educational research: Toward a complex systems conceptual framework of learning. *Educational Psychologist, 51,* 210–218. http://dx.doi.org/10.1080/00461520.2016.1166963

James, C. (2011, September). *Theory of change review: A report commissioned by Comic Relief*. London, UK: Comic Relief. Retrieved from http://www.actknowledge.org/resources/documents/James_ToC.pdf

Jiang, Y., Granja, M. R., & Koball, H. (2017). *Basic facts about low-income children: Children under 18 years, 2015*. New York, NY: National Center for Children in Poverty, Columbia University Mailman School of Public Health. Retrieved from http://www.nccp.org/publications/pdf/text_1170.pdf

Jones, H. (2011). *Taking responsibility for complexity: How implementation can achieve results in the face of complex problems*. (Working Paper 330). London, UK: Overseas Development Institute. Retrieved from https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/6485.pdf

Kauffman, J. M. (2009). Attributions of malice to special education policy and practice. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities: Vol. 22. Policy and practice* (pp. 33–66). Bingley, UK: Emerald.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., . . . Dunlop Velez, E. (2016). *The condition of education 2016* (NCES 2016–144). Washington, DC: United States Department of Education, National Center for Education Statistics.

King, J. B. Jr. (2017, January 17). *A dispatch from the outgoing education secretary: America has the right to a great public education*. Bethesda, MD: Education Week. Retrieved from https://www.edweek.org/ew/articles/2017/01/18/a-dispatch-from-the-outgoing-us-education.html

Kirschner, P. A. (2014). When a paradigm becomes a paradogma. *Journal of Computer Assisted Learning, 30,* 297–299. http://dx.doi.org/10.1111/jcal.12074

Klein, J. T. (1990). *Interdisciplinarity: History, theory, & practice*. Detroit, MI: Wayne State University Press.

LaBoskey, V. K. (1998). The danger of false dichotomies. *Teacher Education Quarterly, 25,* 39–43.

Lamy, C. E. (2013). *American children in chronic poverty: Complex risks, benefit-cost analyses, and untangling the knot*. Lanham, MD: Lexington Books/New York, NY: Crown Publishers.

Lauer, J. M. (1984). Composition studies: Dappled discipline. *Rhetoric Review, 3,* 20–29. http://dx.doi.org/10.1080/07350198409359074

Lyall, C., & Fletcher, I. (2013). Experiments in interdisciplinary capacity-building: The successes and challenges of large-scale interdisciplinary investments. *Science & Public Policy, 40,* 1–7. http://dx.doi.org/10.1093/scipol/scs113

McCarty, T. L., Mancevice, N., Lemire, S., & O'Neil, H. F., Jr. (2017). Introduction: Education research for a new century. A renewed vision of interdisciplinarity. *American Educational Research Journal, 54,* 5S–22S. http://dx.doi.org/10.3102/0002831216687340

Miller, T. R., Baird, T. D., Littlefield, C. M., Kofinas, G., Chapin, F., III, & Redman, C. L. (2008). Epistemological pluralism: Reorganizing interdisciplinary research. *Ecology and Society, 13,* Article 46. Retrieved from http://www.ecologyandsociety.org/vol13/iss2/art46/

Moran, J. (2010). *Interdisciplinarity* (2nd ed.). New York, NY: Taylor & Francis.

Morris, P. A., & Reardon, S. F. (2017). Moving education science forward by leaps and bounds: The need for interdisciplinary approaches to improving children's educational trajectories. *Journal of Research on Educational Effectiveness, 10,* 1–6. http://dx.doi.org/10.1080/19345747.2016.1254466

National Academy of Sciences. (2005). *Facilitating interdisciplinary research*. Washington, DC: National Academies Press.

Nichols, S. L. (2016, May 23). Spreading the word about good research is a 21st century imperative. *Psychology Today, APA Division 15*. Retrieved from https://www.psychologytoday.com/blog/psyched/201605/education-research-and-the-media

Ostrander, R. R. (2015). School funding: Inequality in district funding and the disparate impact on urban migrant school children. *Brigham Young University Education and Law Journal, 1,* 271–295. Retrieved from https://digitalcommons.law.byu.edu/cgi/viewcontent.cgi?article=1362&context=elj

Pelka, F. (2012). *What we have done: An oral history of the disability rights movement*. Amherst, MA: Massachusetts University Press.

Pellmar, T. C., & Eisenbery, L. (Eds.). (2000). *Bridging disciplines in the brain, behavioral, and clinical sciences*. Washington, DC: National Academies Press. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK44872/

Prior, P. (2006). A sociocultural theory of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54–66). New York, NY: Guilford Press.

Putnam, R. D. (2015). *Our kids: The American dream in crisis*. New York, NY: Simon & Schuster.

Resnick, L. B. (1987). Constructing knowledge in school. In L. S. Liben (Ed.), *Development and learning: Conflict or congruence?* (pp. 19–50). Hillsdale, NJ: Erlbaum.

Schmidt, A. H., Robbins, A. S. T., Combs, J. K., Freeburg, A., Jesperson, R. G., Rogers, H. S., . . . Wheat, E. (2012). A new model for training graduate students to conduct interdisciplinary, interorganizational, and international research. *Bioscience, 62,* 296–304. http://dx.doi.org/10.1525/bio.2012.62.3.11

Schwartz, S. J., Lilienfeld, S. O., Meca, A., & Sauvigné, K. C. (2016). The role of neuroscience within psychology: A call for inclusiveness over exclusiveness. *American Psychologist, 71,* 52–70. http://dx.doi.org/10.1037/a0039678

Slavicek, G. (2012). Interdisciplinary: A historical reflection. *International Journal of Humanities and Social Science, 2,* 107–113.

Sparks, S. D. (2017). Commentary: Study of i3 flags issues for school innovators. *Education Week, 36,* 9–11. Retrieved from https://www.edweek.org/ew/toc/2017/01/25/index.html

Spatig-Amerikaner, A. (2012). *Unequal education: Federal loophole enables lower spending on students of color*. Washington, DC: Center for American Progress. Retrieved from https://www.americanprogress.org/wp-content/uploads/2012/08/UnequalEduation.pdf

Staats, A. W. (2005). A road to, and philosophy of, unification. In R. J. Sternberg (Ed.), *Unity in psychology* (pp. 159–177). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10847-010

Staats, A. W. (2016). Neurological exclusiveness or unified science inclusiveness: Comment on Schwartz et al. *American Psychologist, 71,* 894–895, 2016. http://dx.doi.org/10.1037/amp0000032

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution.

Tannen, D. (1998). *The argument culture: Stopping America's war of words*. New York, NY: Random House.

Tseng, V. (2012). Social policy report: The uses of research in policy and practice. *Sharing Child and Youth Development Knowledge, 26,* 1–16.

United States Department of Health and Human Services, National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Reading Panel. (1999). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and implications for reading instruction*. Washington, DC: Author.

Vealey, K. P., & Rivers, N. A. (2014). Dappled discipline at thirty: An interview with Janice M. *Lauer. Rhetoric Review, 33,* 165–180.

Vogel, I. (2012, April). *Review of the use of 'theory of change' in international development*. London, UK: UK Department of International Development. Retrieved from http://www.theoryofchange.org/pdf/DFID_ToC_Review_VogelV7.pdf

Wells, A. (2009). "Our children's burden:" A history of federal education policies that ask (now require) our public schools to solve social inequality. In M. A. Rebell & J. R. Wolff (Eds.), *NCLB at the crossroads* (pp. 1–42). New York, NY: Teachers College Press.

Winzer, M. A. (2009). *From integration to inclusion: A history of special education in the 20th century*. Washington, DC: Gallaudet Press.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330,* 686–688. http://dx.doi.org/10.1126/science.1193147

Received October 27, 2017
Revision received December 20, 2017
Accepted December 20, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 174–179

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000195

Educational Psychology's Past and Future Contributions to the Science of Learning, Science of Instruction, and Science of Assessment

Richard E. Mayer
University of California, Santa Barbara

Patricia Alexander (2018) provides a thought-provoking analysis of the past and future of educational psychology. Based on the themes in Alexander's paper, the present paper explores the past and future of educational psychology's contributions to: (a) the science of learning, corresponding to Alexander's theme of "a focus on learning as a core construct"; (b) the science of instruction, corresponding to Alexander's theme of "a search for evidence-based approaches and practices that work"; and (c) the science of assessment, corresponding to Alexander's theme of "an investment in measurement and an appreciation of human variability." Educational psychology remains poised to make continued contributions to theory and practice.

Keywords: science of learning, science of instruction, science of assessment

Patricia Alexander (2018) provides a thought-provoking analysis of the past and future of educational psychology. Based on the themes in Alexander's paper, the present paper explores the past and future of educational psychology's contributions to: (a) the science of learning, corresponding to Alexander's theme of "a focus on learning as a core construct"; (b) the science of instruction, corresponding to Alexander's theme of "a search for evidence-based approaches and practices that work"; and (c) the science of assessment, corresponding to Alexander's theme of "an investment in measurement and an appreciation of human variability." More specifically, the science of learning refers to the scientific study of how people learn, the science of instruction refers to the scientific study of how to help people learn, and the science of assessment is the scientific study of how to determine what people know.

The relations among these topics is reciprocal, with educational psychology as the link among them. First, educational psychology is a linking science involved in applying the science of learning to educational practice by creating a science of instruction. As eloquently noted by William James (1899/1958, p. 22) in his classic little book, *Talks to Teachers*: "You make a great, a very great mistake, if you think that psychology, being the science of the mind's laws, is something from which you can deduce definite programs and schemes and methods of instruction for immediate classroom use." In short, educational psychology is the link between learning and instruction by creating ways to help people learn based on research-based learning theory and testing them to create a science of instruction. At the same time, educational psychology is a linking science involved in challenging the science of learning to develop theories about academic learning that are relevant to the practical needs of education. In support of this reciprocal relation, educational psychology helps create what can be called a *two-way street* between psychology and education (Mayer, 1992).

When assessment is added to the mix, educational psychology is the link between assessment and instruction, by helping specify learning objectives and learning outcomes in terms of changes in specific knowledge, skills, and beliefs and by helping describe the characteristics of individual learners in terms that are relevant to instruction (e.g., the nature of existing knowledge, skills, and beliefs). Similarly, educational psychology is the link between assessment and learning, by helping specify the cognitive, metacognitive, and motivational processes during learning and by helping specify what is learned in terms of changes in the learner's knowledge, skills, and beliefs. In short, educational psychology is the linking science involved in the assessment of learning outcomes and learning processes.

The time course of events runs from instruction to learning to assessment but it also is iterative (Mayer, 2011). Instruction—the first event—combines with the characteristics of the learner to produce learning—the second event—in the form of cognitive processing and learning outcomes (e.g., knowledge, skills, and beliefs), which can be assessed—the third event—in ways that guide subsequent instruction. In short, three of the key themes discussed by Alexander form a system, with educational psychology making it work. Because of the multiple disciplines involved in this system—including psychologists interested in learning, educators interested in instruction, and statisticians interested in assessment—work in educational psychology reflects "interdisciplinary and cross-disciplinary inquiry," which is another one of Alexander's (2018) themes. Because the central focus is on cognitive processes during learning and learning outcomes, work in educational psychology depends on what Alexander (2018) describes as "the psychologizing of education," which is the last of her five themes.

This article was published Online First March 13, 2017.

Preparation of this paper was supported by grant N00014162046 from the Office of Naval Research.

Correspondence concerning this article should be addressed to Richard E. Mayer, Department of Psychological and Brain Sciences, University of California, 3841 Psychology East, Santa Barbara, CA 93106. E-mail: mayer@psych.ucsb.edu

In short, educational psychology stands as a linking science between psychology and education, which enables and enriches the science of learning, science of instruction, and science of assessment.

Educational Psychology's Past and Future Contributions to the Science of Learning

The science of learning—figuring out how people learn—was one of the original tasks of psychology and education, and continues as a central issue today. Since its founding at the start of 20th century, educational psychology has contributed to the science of learning (a) by shifting the focus from behaviorist to cognitive conceptions of learning, (b) by shifting from general theories of learning to specialized theories of learning in subject areas, and (c) by shifting the focus from learning behavior (i.e., purely behavioral measures) to learning strategies (i.e., measures of cognitive processing during learning).

First, as chronicled by Mayer (1992, 2001), conceptions of learning have progressed from viewing learning as response strengthening to viewing learning as information acquisition to viewing learning as knowledge construction. The first half of the 20th century was dominated by the behaviorist-inspired view of learning as the strengthening and weakening of stimulus–response associations based on rewards and punishments, gleaned largely from research on rats running in mazes or pressing keys in a Skinner box. This view was not adequate to address the challenges of explaining learning in natural contexts, including learning in schools. Instead of building a science of learning based on how lab animals learned in contrived lab tasks, by midcentury pressure mounted to understand how people learn academic material—led by educational psychologists. Educational psychologists, inspired by the practical challenges of education, helped lead the cognitive revolution (Mayer, 2014a) that blossomed in the second half of the 20th century and eventually led to constructivist conceptions of learning—the idea that people are active sense makers who build coherent mental representations by combining aspects of what is presented with what they already know. In short, the single biggest development in the science of learning—the shift from behaviorist to cognitive views of how learning works—was instigated by educational psychology's demands for a theory of learning that is relevant to academic learning.

Second, the science of learning has progressed from attempting to build a general theory of learning to building specific theories of learning for each subject area, which can be called *psychologies of subject matter* (Mayer, 2004). The first half of the 20th century was dominated by competing attempts to build general theories of learning, culminating in Hull's (1943) distillation of the principles of learning into a set of equations based mainly on animal research. Although the learning principles might have applied to certain contrived learning situations with lab animals, they did not seem ideal for application in the world of student learning in schools. By midcentury general theories of learning had run their course and the science of learning probably would have collapsed had it not been for calls for more educationally relevant theories of learning, spearheaded by educational psychologists. The result has been the development of psychologies of subject matter—such as theories of how students learn to read, learn to write, learn mathematics, learn science, learn history, or learn a second language—which represent a unique and monumental contribution of educational psychology to the science of learning that is still strong today (Mayer, 2008; Mayer & Alexander, 2018). Examples include pinpointing the role of the learner's prerequisite knowledge such as phonological awareness in learning to read, number sense in learning arithmetic, and preconceptions in science learning (Mayer, 2008).

Third, learning theories have progressed from a focus on the learner's physical behavior during learning to a focus on the learner's cognitive processing during learning (Mayer, 2009, 2011). For example, in his classic book, *Animal Intelligence*, Thorndike (1911/1965) carefully described the behavior of cats and dogs as they learned how to escape from his puzzle box and described learning outcomes in terms of the strength of each response. Although focusing on learning behavior was useful for developing learning theory in the first half of the 20th century, something more was needed when the research venue shifted to learning of academic content in schools. Educational psychologists were instrumental in helping develop information processing models of how students learn academic content, including selecting relevant information from a lesson, mentally organizing it into a coherent cognitive structure, and integrating it with relevant prior knowledge (Mayer, 2009, 2011). This work lead to research on *learning strategies*–cognitive processing during learning intended to enhance learning—which represents another unique and monumental contribution of educational psychology that is still strong today (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Fiorella & Mayer, 2015). For example, some promising learning strategies that warrant further work include learning by self-testing, learning by self-explaining, learning by teaching others, learning by enacting, learning by summarizing, learning by mapping, learning by drawing, and learning by imagining (Fiorella & Mayer, 2015).

The future involves overcoming some of the limitations of the cognitive revolution by incorporating motivation, metacognition, affect, and brain science into theories of academic learning. First, motivation—an internal state that initiates and maintains goal-directed behavior—is widely recognized as an essential ingredient in academic learning that is reflected in a collection of cognitive theories of academic motivation (Wentzel & Wigfield, 2016). For example, the beliefs about learning that students bring to the learning situation can affect the nature of their learning process. Integrating motivation into the science of learning represents an important continuing goal of educational psychology.

Second, metacognition—which includes awareness and control of one's learning process—is also widely recognized as an essential ingredient in academic learning (Dunlosky & Metcalfe, 2009; Mayer, 2011). Work on learners' judgments of learning, confidence in performing, and control of cognitive processing during learning represent emerging contributions to the science of learning. An important advance is training of specific cognitive processes involved in executive function (Banich, 2009; Miyake et al., 2000), such as shifting (i.e., switching from one task to another), updating (i.e., keeping track of multiple events), and inhibition (i.e., not attending to irrelevant features). Integrating metacognition, including specific cognitive processes involved in executive function, into the science of learning can continue to strengthen learning theories in the future.

Third, affect—which refers to the experience of emotion—is gaining increasing recognition as an essential ingredient in academic learning, sometimes linked to theories of motivation via constructs such as interest or value (Wentzel & Wigfield, 2016), sometimes linked to theories of instruction via constructs such as emotional design (Mayer & Estrella, 2014), and sometimes linked to theories of metacognition based on how students react to impasses in problem solving (Kilpatrick, Swafford, & Findell, 2001). In the future, the science of learning would benefit from a better understanding of how to integrate *cold* theories (i.e., classic cognitive theories) and *hot* theories (i.e., theories involving affect).

Although the database in educational neuroscience has grown greatly over the past two decades (Battro, Fischer, & Lena, 2008; Blakemore & Frith, 2005; Byrnes, 2001; Mareschal, Butterworth, & Tolmie, 2014; Posner & Rothbart, 2007; Sousa, 2011), there is still consensus that brain research has not yet had a significant impact on education or educational psychology (Bowers, 2016; Bruer, 2014; Mayer, 2016). In the future, it would be useful to build better connections between neuroscience and psychology, with the goal of developing a theory of learning that is relevant to education.

Finally, continuing work is needed to overcome some of the limitations of the cognitive revolution by including social, cultural, evolutionary, and situational aspects of learning.

Educational Psychology's Past and Future Contributions to the Science of Instruction

Educational psychology contributed to the science of instruction (a) by amassing a substantial research base pinpointing instructional methods that produce deep learning (i.e., instructional design) and (b) by amassing an emerging research base pinpointing training of learning strategies that produce deep learning (i.e., cognitive process instruction). For example, in Mayer and Alexander's (2018) *Handbook of Research on Learning and Instruction*, chapters summarize research on effective instructional methods including instruction based on feedback, worked-out examples, cooperative learning, inquiry, discussion, tutoring, visualization, computer simulations, and interactive learning technologies. In his monumental meta-analysis of 800 meta-analyses related to academic achievement, Hattie (2009) identified instructional techniques that have been shown to improve learning by at least 0.4 standard deviations, which he considers an educationally significant effect. The list of effective instruction methods includes feedback, worked examples, reciprocal teaching, cooperative learning, direct instruction, peer tutoring, spaced practice, and many more. Another, more specialized example is a set of research-based principles for designing multimedia instructional messages (Mayer, 2009, 2014b). When it comes to training of learning strategies, some effective learning strategies include summarizing, mapping, drawing, imagining, self-explaining, self-testing, teaching, and enacting (Fiorella & Mayer, 2015). Overall, during the past 30 years in particular, educational psychology has developed a sizable research base that supports research-based principles for instruction and training of learning strategies. An important goal for the future is to determine the boundary conditions for each research-based principle of instruction, including when it works, for whom it works, and for which kind of learning objectives it works, as well as to determine how it plays out in educational contexts and with new media.

Educational Psychology's Past and Future Contributions to the Science of Assessment

Educational psychology contributed to the science of assessment by developing techniques for assessing (a) types of knowledge and skills (i.e., learning outcomes), (b) types of cognitive processing during learning (i.e., learning processes), and (c) types of learners (i.e., learner characteristics).

First, cognitive testing has been a core component of educational psychology from its inception, epitomized by Thorndike's work on measurement of individual differences, which included developing standardized tests of school achievement in subjects such as reading, arithmetic, and handwriting; developing a standardized college admission test; being part of a team that developed the first large-scale selection tests for the U.S. Army in World War I; and being part of a team that professionalized psychological testing by founding the Psychological Corporation in 1921 (Mayer, 2003). Thorndike (1918, p. 16) set the tone for psychological testing with his famous quote: "Whatever exists at all exists in some amount." When it comes to cognitive testing, educational psychologists such as Thorndike (Mayer, 2003) and Binet (Wolf, 1973) offered a shift from viewing intellectual ability as a mental factor—which was the dominant view the first half of the 20th century—to viewing intellectual ability as based on knowledge acquisition. Bloom's taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) represents an important step in building a taxonomy of the kinds of learning outcomes that could be subjected to targeted testing. Today, there is growing consensus that cognitive performance depends on what the learner knows, so the focus of cognitive assessment should be on determining the learner's existing knowledge, skills, and beliefs (Anderson et al., 2001; Pellegrino, Chudowsky, & Glaser, 2001). An important contribution of educational psychology has been on analyzing and measuring types of knowledge, such as factual, conceptual, procedural, and metacognitive knowledge (Anderson et al., 2001) or facts, concepts, procedures, strategies, and beliefs (Mayer, 2011).

In the future, instead of high-stakes summative testing conducted outside the learning environment that dominates educational assessment today, educational psychologists should lead the shift to low-stakes formative assessment that is embedded within the natural course of learning. The goal is to provide a continuous and unobtrusive monitoring of learning so that both students and teachers can see individual growth in knowledge, which Hattie (2009) refers to as *visible learning*. Computer-based technology is likely to play a useful role in helping monitor each student's growth in knowledge, analogous to the use of self-monitoring devices in fitness that provide a continuous reading of miles walked, steps climbed, heart rate, and the like. Real-time monitoring of each learner's knowledge, motivation, affect, and metacognition can also help instructors adapt their instruction, so a focus on building feedback that leads to more effective adaptive instruction is an important related goal for the future. For example, Shute and Ventura (2013) have shown how learning assessments can be embedded within computer games to create *stealth assessment*; that is, assessments that appear to be part of computer-based activities to learners.

Second, educational psychology has been at the forefront of assessing cognitive processing during learning using a variety of techniques ranging from self-report surveys to thinking aloud protocols to data mining of button presses in online learning to physiological measures. At a gross level, such processes can be characterized as selecting (i.e., attending to relevant incoming information), organizing (i.e., constructing coherent structures), and integrating (i.e., connecting incoming information with relevant prior knowledge; Mayer, 2009, 2011). At a more domain-specific level, each kind of academic task can be analyzed into subprocesses such as recognizing phonemes, decoding words, developing fluency, and accessing word meaning in reading; using prior knowledge, using prose structure, making inferences, and comprehension monitoring in reading comprehension; planning, translating, and reviewing in writing; or problem translation, problem integration, solution planning, solution monitoring, and solution execution in mathematical problem solving (Mayer, 2008). In short, an important contribution of educational psychology has involved assessing the learner's cognitive processing during learning.

In the future, physiological measures, particularly measures of brain activity such as fMRI and EEG, may prove helpful in supplementing self-reports of cognitive activity during learning. Similarly, another way to supplement self-report measures of cognitive activity during learning involves computer-based technologies that can record relevant activities during learning (such as button presses, pen strokes, or eye movements). Refinement of online measures of affect during learning represents another important future direction for assessment in the future.

Third, educational psychology has been instrumental in highlighting the role of individual differences in learning and instruction. Importantly, research has called into question the idea that instruction should be adapted to each student's learning style, such as using verbal methods for verbal learners and visual methods for visual learners (Massa & Mayer, 2006; Pashler, McDaniel, Rohrer, & Bjork, 2008). Although research shows that a focus on individual differences in learning styles may not be productive (Holmes, 2016; Hunt, 2011), there is solid support for the idea that the single most important individual differences dimension for educational practice is prior knowledge (Mayer, 2008, 2009, 2011). The essential role of prior knowledge in meaningful learning is documented in research showing that learning involves assimilating incoming information with existing schemas, so that meaningful learning is problematic when students lack appropriate schemas (Ausubel, 1960; Bartlett, 1932; Sweller, Ayres, & Kalyuga, 2011). For example, Kalyuga (2014) has documented what he calls the *expertise reversal effect* in which instructional methods that are optimal for less knowledgeable learners (such as heavily guided instruction) may not be optimal for more knowledgeable learners, and vice versa. There is also emerging evidence that the motivational beliefs that students bring to the learning situation can affect learning, which has resulted in validated surveys of various kinds of motivational beliefs (Wentzel & Wigfield, 2016). In short, educational psychology has been at the forefront in documenting the importance of individual differences in prior knowledge and motivation.

In the future, an important goal for educational psychology will be to devise more efficient and valid ways to assess individual differences in prior knowledge, motivation, and metacognition that can be used to guide instruction.

Resolving Challenges in Educational Psychology

Our field has been beset by a collection of challenges in the past—including searching for the perfect "ism", pledging allegiance to methodology-centered research (sometimes called waging the paradigm wars), taking excursions into the post-truth era, and neglecting the role of replication. These challenges have sometimes slowed progress in our field so it is worthwhile to resolve them to insure a more productive path in the future.

First, some educational researchers have devoted much energy to an unproductive search for the right "ism"—ranging from cognitivism to constructivism to constructionism to social constructionism to radical social constructivism and so on (Phillips & Burbules, 2000). We have learned that our field is set back when theory building is no longer based on evidence gleaned from scientifically sound studies but rather becomes an exercise in building untestable doctrine to which educational practices must adhere. From my vantage point, it appears that a bright future depends on our commitment to taking a scientific approach, in which educational practice is based on research evidence and research-based theory, rather than a doctrine-based approach, in which educational practice must conform to the slogans of popular "isms".

Second, considering the trend toward methodology-centered research, our field has suffered from an unproductive commitment by some researchers to define research in terms of research methods—for example, conducting observational studies versus experiments or collecting qualitative versus quantitative data—rather than research goals. We have learned that our field is set back when researchers are trained to use one methodological approach regardless of the research question being addressed, because as eloquently noted by Shavelson and Towne (2002. p. 63), "the method used to conduct scientific research must fit the question posed, and the investigator must competently implement the method." For example, when the goal is to draw causal claims about whether an instructional intervention is effective, experimental methods with quantitative measures are called for (Phye, Robinson, & Levin, 2005). Thus, it would be a disservice to dismantle training in experimental and quantitative research methods in schools of education, creating researchers who are committed solely to using observational and qualitative techniques. From my vantage point, it appears that the future is bright to the extent that educational researchers are committed to using research methods that match their research goals.

Third, the broader field of educational research has sometimes been threatened by those who are so committed to their personal beliefs and political opinions that the role of research evidence is diminished—leading to what can be called a *post-truth era*. In response to this unproductive commitment to opinions and beliefs rather than evidence, Shavelson and Towne (2002, p. 25) state "we reject the postmodernist school of thought when it posits that social science research can never generate objective or trustworthy knowledge." From my vantage point, what holds our field together today and in the future is our mutual commitment to basing our arguments on evidence gleaned from methodologically sound research.

Fourth, our field has sometimes been hesitant to recognize the value of replication studies, with some journals reluctant to publish papers that replicate previous work. In response to this unproductive stance, Shavelson and Towne (2002, p. 70) list "replicate and generalize across studies" as one of the six fundamental principles for scientific research in education. For example, recent conversations about the "crisis of replication" in psychological research call into question whether some well-known effects can be replicated (Pashler & Wagenmakers, 2012, p. 528; Stroebe & Strack, 2014, p. 59). As our field comes to increasingly value the place of meta-analysis in resolving questions about instructional interventions (Hattie, 2009), for example, we are learning to value the role of replication studies. From my vantage point, the future is bright to the extent that conclusions are drawn from a substantial research base of replication studies rather than from a single study.

Conclusion

In this paper, I attempted to synthesize three of the key themes in Alexander's (2018) stimulating analysis of the past and future of educational psychology—concerning the contributions of educational psychology to the science of learning, science of instruction, and science of assessment. Given the impossibility of documenting every contribution of educational psychology over its first 100+ years, I focused on exemplary contributions that I consider major and unique based on my perspective of over 40 years of research work in applying the science of learning to education.

One constant force in educational psychology is our reliance on science, including rigorous scientific research methods, as exemplified by the careful experiments of the world's first educational psychologist, E. L. Thorndike (Mayer, 2003). Our field's commitment to science is reflected in Thorndike's (1906, p. 206) call for educators to "direct their work by scientific spirit and methods" and "direct their choices of methods by the results of scientific investigation rather than general opinion." In the future, education is likely to continue to be under assault by those who prefer to base instructional decisions on opinions, fads, ideology, or advocacy, and psychology is likely to continue to be under assault by those who would prefer to reduce psychology to mathematics, biology, and chemistry. In my opinion, the future of educational psychology will be bright to the extent that we repel these assaults and remain true to our constant commitment to scientific research methods for understanding the workings of learning, instruction, and assessment.

References

Alexander, P. A. (2018). Past as prologue: Educational psychology's legacy and progeny. *Journal of Educational Psychology, 110,* 147–162. http://dx.doi.org/10.1037/edu0000200

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., . . . Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York, NY: Longman.

Ausubel, D. P. (1960). *Educational psychology: A cognitive view.* New York, NY: Holt, Rinehart, and Winston.

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18,* 89–94.

Bartlett, F. (1932). *Remembering.* New York, NY: Cambridge University Press.

Battro, A. M., Fischer, K. W., & Lena, P. J. (Eds.). (2008). *The educated brain.* New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511489907

Blakemore, S.-J., & Frith, U. (2005). *The learning brain.* Malden, MA: Blackwell.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives. Handbook I: Cognitive domain.* New York, NY: David McKay.

Bowers, J. S. (2016). The practical and principled problems with educational neuroscience. *Psychological Review, 123,* 600–612. http://dx.doi.org/10.1037/rev0000025

Bruer, J. T. (2014). Afterword. In D. Mareschal, B. Butterworth, & A. Tolmie (Eds.), *Educational neuroscience* (pp. 349–363). Malden, MA: Wiley Blackwell.

Byrnes, J. P. (2001). *Minds, brains, and learning.* New York, NY: Guilford Press.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition.* Thousand Oaks, CA: SAGE.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. http://dx.doi.org/10.1177/1529100612453266

Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity: Eight learning strategies that promote understanding.* New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781107707085

Hattie, J. (2009). *Visible learning.* New York, NY: Routledge.

Holmes, J. D. (2016). *Great myths of education and learning.* Malden, MA: Wiley. http://dx.doi.org/10.1002/9781118760499

Hull, C. (1943). *Principles of behavior.* New York, NY: Appleton-Century-Crofts.

Hunt, E. (2011). *Human intelligence.* New York, NY: Cambridge University Press.

James, W. (1958). *Talks to teachers.* New York, NY: Norton. (Original work published 1899)

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 576–597). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.028

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Mareschal, D., Butterworth, B., & Tolmie, A. (Eds.). (2014). *Educational neuroscience.* Malden, MA: Wiley Blackwell.

Massa, L. J., & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences, 16,* 321–335. http://dx.doi.org/10.1016/j.lindif.2006.10.001

Mayer, R. E. (1992). Cognition and instruction: On their historic meeting within educational psychology. *Journal of Educational Psychology, 84,* 405–412. http://dx.doi.org/10.1037/0022-0663.84.4.405

Mayer, R. E. (2001). Changing conceptions of learning: A century of progress in the scientific study of education. In L. Corno (Ed.), *Education across a century: The centennial volume. Hundredth yearbook of the National Society for the Study of Education* (pp. 34–75). Chicago, IL: National Society for the Study of Education.

Mayer, R. E. (2003). E. L. Thorndike's enduring contributions to educational psychology. In B. J. Zimmerman & D. H. Schunk (Eds.), *Educational psychology: A century of contributions* (pp. 113–154). New York, NY: Routledge.

Mayer, R. E. (2004). Teaching of subject matter. In S. T. Fiske (Ed.), *Annual review of psychology* (Vol. 55, pp. 715–744). Palo Alto, CA: Annual Reviews.

Mayer, R. E. (2008). *Learning and instruction* (2nd ed.). Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511811678

Mayer, R. E. (2011). *Applying the science of learning*. Upper Saddle River, NJ: Pearson.

Mayer, R. E. (2014a). Cognitive revolution and information processing perspective. In D. C. Phillips (Ed.), *Encyclopedia of educational theory and philosophy* (pp. 142–149). Thousand Oaks, CA: SAGE.

Mayer, R. E. (Ed.). (2014b). *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369

Mayer, R. E. (2016). How can brain research inform academic learning and instruction? *Educational Psychology Review*. Advance online publication. http://dx.doi.org/10.1007/s10648-016-9391-1

Mayer, R. E., & Alexander, P. A. (Eds.). (2018). *Handbook of research on learning and instruction* (2nd ed.). New York, NY: Routledge.

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33,* 12–18. http://dx.doi.org/10.1016/j.learninstruc.2014.02.004

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100. http://dx.doi.org/10.1006/cogp.1999.0734

Pashler, H., McDaniel, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological Science in the Public Interest, 9,* 105–119. http://dx.doi.org/10.1111/j.1539-6053.2009.01038.x

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis in confidence? *Perspectives on Psychological Science, 7,* 528–530. http://dx.doi.org/10.1177/1745691612465253

Pellegrino, J. G., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowman & Littlefield.

Phye, G. D., Robinson, D. H., & Levin, J. (Eds.). (2005). *Empirical methods for evaluating educational interventions*. San Diego, CA: Elsevier Academic Press.

Posner, M. I., & Rothbart, M. K. (2007). *Educating the human brain*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/11519-000

Shavelson, R. J., & Towne, L. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Shute, V., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. Cambridge, MA: MIT Press.

Sousa, D. A. (2011). *Educational neuroscience*. Thousand Oaks, CA: Corwin. http://dx.doi.org/10.4135/9781483387734

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9,* 59–71. http://dx.doi.org/10.1177/1745691613514450

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4419-8126-4

Thorndike, E. L. (1906). *Principles of learning based on psychology*. New York, NY: Seiler. http://dx.doi.org/10.1037/11487-000

Thorndike, E. L. (1918). The nature, purposes, and general methods of measurement of educational products. In G. M. Whipple (Ed.), *17th Yearbook of the National Society for the Study of Education, Part 2: The measurement of educational products* (pp. 16–24). Bloomington, IL: Public School Publishing Company.

Thorndike, E. L. (1965). *Animal intelligence*. New York, NY: Hafner. (Original work published 1911)

Wentzel, K. R., & Wigfield, A. (2016). *Handbook of academic motivation at school* (2nd ed.). New York, NY: Routledge.

Wolf, T. H. (1973). *Alfred Binet*. Chicago, IL: University of Chicago Press.

Received November 3, 2016
Revision received January 24, 2017
Accepted January 24, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 180–202

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000216

Does Writing System Influence the Associations Between Phonological Awareness, Morphological Awareness, and Reading? A Meta-Analysis

Yufang Ruan
Beijing Normal University

George K. Georgiou
University of Alberta

Shuang Song, Yixun Li, and Hua Shu
Beijing Normal University

Differences in how writing systems represent language raise important questions about the extent to which the role of linguistic skills such as phonological awareness (PA) and morphological awareness (MA) in reading is universal. In this meta-analysis, the authors examined the relationship between PA, MA, and reading (accuracy, fluency, and comprehension) in 2 languages (English and Chinese) representing different writing systems (alphabetic and logographic). A random-effects model analysis of data from 64 studies with native speakers of each language revealed significant correlations between PA, MA, and all reading outcomes in both languages. The correlations remained significant even after controlling for each other's effect on reading. However, PA was a stronger correlate of reading in English than in Chinese. MA was as good a correlate of reading in English as in Chinese (except for comprehension, where it was better). In addition, complex PA tasks in English and production/compounding MA tasks in Chinese produced significantly larger correlations with reading accuracy. Taken together, the findings of this meta-analysis suggest that PA and MA are significant correlates of reading, but their role is influenced by the writing system, the type of reading outcome, and the type of task used to operationalize PA and MA. The implications of these findings are discussed.

Educational Impact and Implications Statement

The authors examined the role of writing system in the relationship between phonological awareness, morphological awareness, and reading. The results of the meta-analysis revealed significant relationships between these linguistic skills and reading in each language, but the strength of the relationships was influenced by the writing system, the type of reading outcome, and the type of task used to operationalize phonological awareness and morphological awareness. These findings help us better understand the linguistic skills that are most important for reading acquisition in different writing systems.

Keywords: phonological awareness, morphological awareness, reading, meta-analysis

Phonology, orthography, and semantics are three of the major lexical constituents that contribute to reading development (e.g., Kamhi & Catts, 2012; Perfetti, Liu, & Tan, 2005). Connectionist models of reading (see e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) have used these language parameters and proposed that word identification involves making connections between orthography and phonology, and between orthography and semantics. Nevertheless, different writing systems differ in the way they represent language in written form (whereas alphabetic systems such as English use a small set of letters to represent sounds, logographic systems such as Chinese use logograms to represent meaning) and in the statistical properties of the orthography-to-phonology and orthography-to-meaning mappings (e.g., Yang, Shu, McCandliss, & Zevin, 2013). This implies that the role of processing skills that represent phonology (e.g., phonological awareness [PA]) and semantics (e.g., morphological awareness [MA]) in reading may also differ across writing systems. Although a few studies have examined this hypothesis (Cho, Chiu, & McBride-Chang, 2011; McBride-

This article was published Online First June 12, 2017.

Yufang Ruan, State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University; George K. Georgiou, Department of Educational Psychology, University of Alberta; Shuang Song, State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University; Yixun Li, Beijing Key Laboratory of Applied Experimental Psychology, School of Psychology, Beijing Normal University; Hua Shu, State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University.

This research was supported by the National Key Basic Research Program of China (2014CB846103), by the Natural Science Foundation of China (31271082, 31671126, 31611130107), and by the Beijing Municipal Science & Technology Commission (Z151100003915122). The authors would like to thank John Kirby and Peter Bowers for their valuable comments on an earlier draft of this manuscript.

Correspondence concerning this article should be addressed to George K. Georgiou, Department of Educational Psychology, 6-102 Education North, University of Alberta, Edmonton-AB, Canada T6G2G5. E-mail: georgiou@ualberta.ca

Chang, Cho, et al., 2005, 2012), to date, no systematic reviews have been conducted. Thus, the purpose of this meta-analysis was to examine if the size of the relationship between PA (the ability to access and manipulate speech sounds in a language), MA (the awareness of morphemic structures of words and the ability to reflect on them), and reading differs between English and Chinese, two languages with distinct linguistic features.

English has been described as a morphophonemic language (Chomsky, 1970; Venezky, 1967). Because it uses an alphabetic script, to read a given word one would need to know how symbols (i.e., graphemes) relate to sounds (i.e., phonemes). However, in English, these grapheme-phoneme correspondences are highly inconsistent: a letter can be pronounced in different ways and a phoneme can be spelled in various ways. To resolve this problem, when two words are pronounced the same but have different meanings (e.g., *to*, *two*, *too*), spelling has evolved, where possible, to separate those meanings with different spellings. As stated by Venezky (1967) five decades ago, "the simple fact is that the present orthography [English] is not merely a letter-to-sound system riddled with imperfections, but instead, a more complex and more regular relationship wherein phoneme and morpheme share leading roles" (p. 77).

These features of English are in contrast to those of Chinese, which has been described as a morphosyllabic language that uses a logographic script (Hanley, 2005; Shu, 2003). The basic graphic unit in Chinese is the character, which corresponds to a monosyllabic morpheme. Characters are made up of a number of strokes that are packed into a square configuration and usually consist of two components: a phonetic radical that gives some clues to the character's pronunciation and a semantic radical that provides information about the meaning of the character. The Chinese characters map onto phonology at the syllabic level, with no parts in a character corresponding to phonological segments like phonemes. Although about 80% of modern Chinese are compound characters containing a phonetic radical, only one fourth of them can be read accurately using the phonetic radical (Chung & Leung, 2008; however, see Shu, Chen, Anderson, Wu, & Xuan, 2003, for a higher estimate).

Because the phonetic information in Chinese characters is encoded at the syllabic level, researchers have argued that the ability to dissect syllables into onsets and rimes should be a significant correlate of Chinese word reading, a hypothesis that has been confirmed in several previous studies (e.g., Ho & Bryant, 1997; McBride-Chang & Ho, 2000; Pan et al., 2011; Shu, Peng, & McBride-Chang, 2008; Zhang et al., 2013). An important role of syllabic awareness or onset/rime awareness in character recognition would also be expected given that Chinese children are introduced to a phonetic alphabet called *Pinyin* (in mainland China) or *Zhuyin Fuhao* (in Taiwan) that is used to assist them in learning new characters. However, Newman, Tardif, Huang, and Shu (2011) showed that phonemic awareness (operationalized with initial, middle, and final phoneme deletion tasks) also predicts Chinese reading, even after controlling for the effects of pinyin knowledge, vocabulary, and syllabic awareness. Taken together, these findings suggest that PA underlies successful reading acquisition in all languages, but the linguistic level that drives its relationship with reading may be the one with the greatest variability at the time of testing (see Siok & Fletcher, 2001, for a similar conclusion).

In contrast to the orthography-to-phonology mapping, the orthography-to-semantics mapping is more reliable in Chinese than in alphabetic orthographies. The semantic radical in Chinese characters provides useful cues to the meaning of a character (e.g., Hanley, 2005). However, because there are about 7,000 morphemes, but only 1,300 syllables in Mandarin Chinese (Chao, 1976), more than five morphemes share the same syllable (Packard, 2000). Hence, a reader must be able to distinguish between homophone characters that share the same syllable (e.g., /yi4/), but with different morphemes (e.g., 义 'meaning,' 易'easy,' 亿 'a hundred million,' 异 'difference,' 益 'benefit,' 艺 'art,' 议 'discuss'). This renders MA (often operationalized in Chinese with homophone awareness tasks) a crucial skill in learning to read Chinese (e.g., Kuo & Anderson, 2006; McBride-Chang, Shu, Zhou, Wat, & Wagner, 2003; Shu, McBride-Chang, Wu, & Liu, 2006; Tong et al., 2011).

Researchers have further argued that because Chinese is a morphosyllabic language, to read, an individual should map the character to a morpheme (Shu, 2003). This should also draw upon one's awareness of morphemes within a word. For example, the meaning of the compound word 大人/da4ren2/adult can be derived from its constituent morphemes, 大/da4/grown and 人/ren2/person. Furthermore, because about 70% of Chinese words are polymorphemic compounds made up of two or more morphemes (Institute of Language Teaching and Research of China, 1986), understanding how morphemes can be legally combined to form a word should play an important role in learning to read Chinese words. Awareness of compound word construction and production in Chinese is also important for vocabulary development, because it helps children to access the meaning of new words based on morphemes they are familiar with (e.g., McBride-Chang et al., 2011; McBride-Chang, Shu, Ng, Meng, & Penney, 2007; Tong, Tong, & McBride, 2017). Given that MA contributes to the learning of new vocabulary (this also applies to other languages and not just Chinese) and that vocabulary and word reading are independent predictors of reading comprehension (e.g., Kendeou, van den Broek, White, & Lynch, 2009; Li, Dronjic, Chen, Li, Cheng, & Wu, in press), we would expect MA to be a particularly strong correlate of reading comprehension. In addition, because MA involves the integration of semantic, phonological, and syntactic information, it mirrors many of the integrative processes involved in reading comprehension (e.g., Kuo & Anderson, 2006; Perfetti, Landi, & Oakhill, 2005). In line with this argument, Chik and colleagues (2012) have shown that MA and morphosyntactic awareness were significant predictors of reading comprehension, even after controlling for the effects of word reading.

There are three levels of MA in Chinese (Li, Anderson, Nagy, & Zhang, 2002; Liu & McBride-Chang, 2010; Shu et al., 2006). The first relates to homophone awareness that has been described above. The second relates to homograph awareness, which requires children to be aware that a single written character (e.g., 草) may represent different morphemes (grass or hasty). Different morphemes contribute to the word's meaning when they are in different compound words (e.g., grass in 草地 lawn or hasty in 草率 cursory). The third relates to the knowledge of the morphemic structure of compound words, which requires awareness of the contribution of the individual morpheme (e.g., 飞 fly and 机 machine) to the meaning of the whole word (e.g., 飞机, airplane). Although several studies have established that MA is a strong

concurrent and longitudinal predictor of Chinese reading (e.g., Liu & McBride-Chang, 2010; McBride-Chang, Cho et al., 2005; Xue, Shu, Li, Li, & Tian, 2013; Yeung et al., 2011), it remains unknown if different levels of MA relate to Chinese reading the same way.

In addition, it remains unclear if the relationship between MA and reading changes over time. Based on phase theories of reading development (e.g., Ehri, 2005; Seymour, 2005) as well as on current practices of teaching reading (e.g., Grade 1–2 teachers in North America put heavy emphasis on phonics that relies on PA; Grade 1–2 teachers in mainland China use Pinyin to introduce new characters with little reference to morphemes), one would expect PA to be more important during the early phases of reading development and MA to be more important during the later phases of reading development.[1] Kuo and Anderson (2006) pointed out that "morphological awareness becomes an increasingly important predictor of measures of reading as children grow older" (p. 161).

Although the few cross-sectional studies in Chinese have confirmed the increasing role of MA in reading over time (e.g., Hu, 2013; Li et al., 2002; Wei et al., 2014; Xue et al., 2013), the few cross-sectional studies in English have either covered the early elementary grades (e.g., Deacon, 2012) or the upper elementary grades (e.g., Nagy, Berninger, & Abbott, 2006; Roman, Kirby, Parrila, Wade-Woolley, & Deacon, 2009), and have provided mixed findings. For example, whereas Nagy et al. (2006) found MA to be a stronger predictor of reading comprehension in older children, Roman et al. (2009) found no age differences when predicting word and nonword reading. Nevertheless, studies with young children in Chinese (e.g., Li, Shu, McBride-Chang, Liu, & Peng, 2012; McBride-Chang et al., 2003; Tong et al., 2011) and English (e.g., Carlisle, 1995; Kirby et al., 2012) have shown that MA is a unique predictor of word reading, even after controlling for the effects of PA. Thus, a meta-analysis is needed to examine if grade level influences the role of MA in reading.

Finally, we do not know if the relationship between MA and reading varies as a function of reading ability status. Despite the findings of studies in both English and Chinese showing that children with dyslexia or specific poor comprehension perform worse than chronological-age controls in different MA tasks (Shu et al., 2006; Siegel, 2008; Tong et al., 2011; Zhang, in press) to our knowledge, no studies have examined the role of reading ability status in the relationship between MA and reading. At the same time, the few studies that examined the role of reading ability status in the relationship between PA and reading have provided mixed findings. For example, McBride-Chang and Manis (1996) reported significantly lower correlations between PA and word reading in the group of poor readers than in the group of good readers, Katzir, Kim, Wolf, Kennedy, Lovett, and Morris (2006) reported no differences between groups, and Savage, Frederickson, Goodwin, Patni, Smith, and Tuersley (2005) reported stronger associations in the group of poor readers.

The Present Study

The purpose of this meta-analysis was to examine if the relationship between PA, MA, and reading (accuracy, fluency, and comprehension) differs between English and Chinese. If the role of PA or MA in reading depends on the linguistic properties of a language, we should observe a stronger relationship between PA and reading in English than in Chinese, and a stronger relationship between MA and reading in Chinese than in English.

The findings of this meta-analysis are expected to make two important contributions to the literature: First, although there are two meta-analyses examining the relationship between PA and reading in English (Scarborough, 1998; Swanson, Trainin, Necoechea, & Hammill, 2003),[2] none of them has examined the relationship between PA and reading fluency. Swanson et al. (2003) further showed that there were no significant differences in the correlations with word reading ($r = .51$) and reading comprehension ($r = .49$). Likewise, Song et al.'s (2016) meta-analysis in Chinese did not examine the association between PA and reading comprehension. No significant differences in the relationship of PA with reading accuracy ($r = .36$) and reading fluency ($r = .39$) were reported. Second, to our knowledge, this is the first meta-analysis of correlational studies examining the relationship of MA with reading in any language.[3] This is important in light of the increased use of MA tasks in research across languages. In their review paper, Nunes and Hatano (2004) suggested that despite the differences between writing systems, MA is important for reading acquisition across languages.

Method

Data Collection and Inclusionary Criteria

The data collection, coding, and inclusionary criteria are summarized in Figure 1. To select the studies for our meta-analysis, we first searched in computerized databases (ERIC, Medline, PsychAPA, PsychInfo, ProQuest, and Google Scholar) for studies published in English from January 1975 to July 2015 using the following descriptors: English, Chinese, China, Hong Kong, Taiwan paired with phon* awareness, phonological processing, MA, reading, and character recognition. Abstracts of peer-reviewed studies, dissertations, and book chapters were subsequently scrutinized. Similar to previous meta-analyses (see Song et al., 2016; Swanson et al., 2003), only studies including both PA and MA measures were considered. This was done to increase our control over possible confounding variables (e.g., age of participants, sampling procedures) associated with different studies and for practical reasons since there are hundreds of studies on PA in English alone.

[1] Had we adopted a different theoretical framework (e.g., overlapping waves; see also Treiman & Kessler, 2014, for integration of multiple patterns framework for spelling development), we should find no developmental differences in the relationship of phonological awareness and morphological awareness with reading. This is because children in kindergarten or Grade 1 have some morphological awareness (e.g., Grigorakis, 2014; Kirby et al., 2012; Li et al., 2012) and can use it together with phonological awareness in word recognition.

[2] Melby-Lervåg, Lyster, and Hulme (2012) also performed a meta-analysis on the relationship between phonemic awareness and word reading but included studies across several languages. The average correlation in their meta-analysis was .57 (95% CI: .54, .59).

[3] However, there are two meta-analyses on the role of morphological awareness instruction in reading ability (see Bowers, Kirby, & Deacon, 2010; Goodwin & Ahn, 2010). Reed (2008) and Carlisle, McBride-Chang, Nagy, and Nunes (2010) also reported significant effects of morphological awareness instruction (particularly for the less able and younger children) in their systematic reviews.

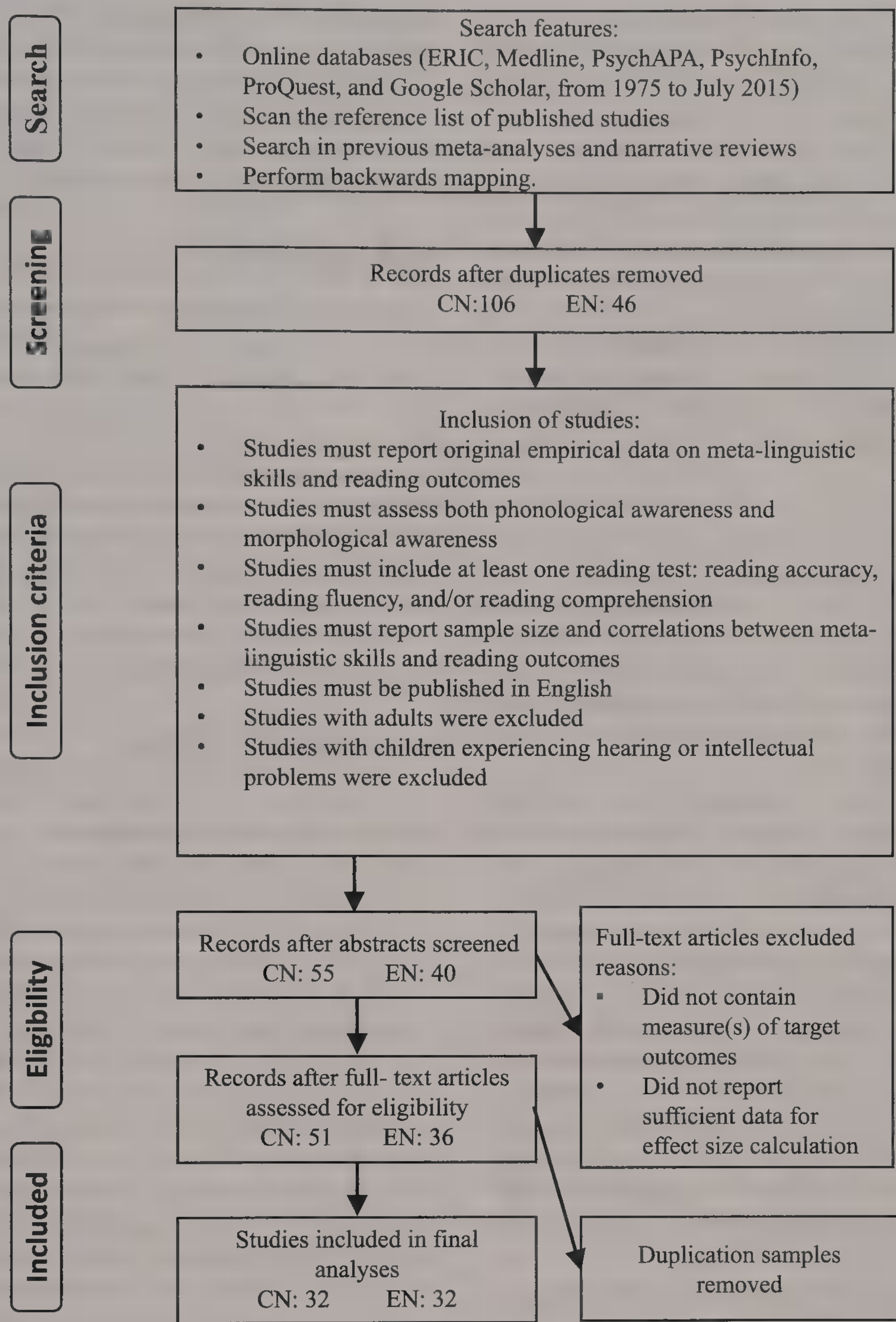

Figure 1. Flow diagram for the search and inclusion of studies. CN = Chinese; EN = English.

Pre-established criteria were used to evaluate the appropriateness of the measures used to assess reading, PA, and MA. Reading accuracy included measures requiring accurate word/character recognition without imposing any time limits. To be considered a measure of reading fluency, the task should require children to read as many words/characters or sentences as possible within a specified time limit. Text reading accuracy or speed was assessed in only two studies in English (no studies in Chinese) and was not considered further. Reading comprehension included measures requiring children to answer questions about a story they read as well as measures requiring children to either provide a missing word that completes the meaning of a sentence/short passage or select the right word among options. For PA, acceptable measures were considered those that involved manipulation of syllables or phonemes of real words/nonwords as in syllable/phoneme deletion/detection test, phoneme blending test, rhyme detection/production test, syllable counting test, and tone detection test (used only in Chinese). Finally, acceptable measures for MA were considered those that involved manipulation (identification, generation) of morphemes in real words or nonwords (found only in English) as in judgment of word relation tests, production of word form tests, and compound structure tests.

To avoid including data from the same study more than once, studies conducted by the same author were further scrutinized. In longitudinal studies, data from the first measurement of each processing skill were coded. In addition, although our studies included native speakers of Chinese and English, for three studies with bilingual children (one with English-Arabic, two with Chinese-English), we only coded data from the children's native language (L1) and ran the analyses with and without these three

studies. Finally, for studies using multiple measures of PA or MA, a set of rules was further established to assist us in coding. For PA, phoneme deletion or tone detection tasks were coded before other types of PA measures because of their complexity as well as predictive value (e.g., Shu et al., 2008). An arithmetic mean of reported *r* values was coded for each sample when more than one test was used to operationalize a type of PA measure (see below for a description of the two types of PA measures). For MA, production tasks were coded before judgment tasks because they are generally more difficult and guessing rates are much lower in them (Deacon, Parrila, & Kirby, 2008; Kirby et al., 2012). Again, an arithmetic mean of reported *r* values was coded, when there were more than one tests used to operationalize a type of MA measure (see below for details).

Moderator Variables

For each study, we coded the following moderators: task type, grade level, and reading ability status. Studies that reported combined scores (e.g., correlations derived from a pooled sample of poor readers and controls) or had limited/no information on the measures they used that prohibited us from obtaining a clear picture of the group in which a certain task could be classified were excluded.

Task type. For PA, phoneme deletion, phoneme blending, phoneme segmentation, phoneme isolation, spoonerism, and tone detection tasks were coded as "complex" and the rest (e.g., syllable deletion/detection task, rhyme detection/production task, onset/rime awareness, and syllable counting) were coded as "simple" PA tasks. For MA, we used three types of classifications: First, we grouped the tasks into two categories, production and judgment. The analysis using this grouping was carried out in both Chinese and English. Second, we grouped the tasks into oral and written groups based on how the morphological tasks were presented to the children. However, because no studies in Chinese had used a written presentation of the MA tasks, this analysis was conducted only with the English studies. Finally, following previous categorizations of MA tasks in Chinese (e.g., Li et al., 2002; Liu & McBride-Chang, 2010; Shu et al., 2006; Tong et al., 2017), we grouped the Chinese tests into three categories: compounding (e.g., "when we see the sun rising in the morning, we call it "sunrise", what would we then call the moon rising in the evening?"); homophone (e.g., 时光 (shi2 guang1, time), 食品 (shi2 pin3, food), 识别 (shi2 bie2, recognize), and 石块 (shi2 kuai4, stone), select which one corresponds to the meaning of (shi 2 bie2, recognize), and homograph (e.g., 草地 (cao3di4, lawn), the children were asked using the target morpheme 草(cao3) to produce two more words, one sharing the same meaning with the target word and one having a different meaning from the target word).

Grade level. Grade level was coded to differentiate between reading development phases. Samples consisting of kindergarten children were coded as "preschooler;" Grade 1 and 2 were coded as "beginning;" Grade 3 and 4 were coded as "intermediate;" and Grade 5 and above (to high school) were coded as "advanced." Studies with adults were excluded from this meta-analysis.

Reading ability status. More than half of the studies in each language used unselected samples of children, which we coded here as "unselected." Control groups in comparison studies or samples clearly described as having no reading problems (or learning disabilities, educational difficulties, developmental disorders) were coded as "normal." Samples including children with dyslexia, poor readers, or at-risk children were coded as "poor." One study with participants experiencing speech disorders and one study with participants who were identified as having a specific language impairment were excluded.

Coder Reliability

All studies were coded twice by the first and the fourth author who received specialized training in meta-analysis. Interrater reliability was calculated for the whole sample of studies. The interrater correlation (Pearson's) for the *r* values (the correlation between PA tasks or MA tasks and reading) was .998 ($p < .001$, agreement rate = 99.3%, $N = 611$). The interrater correlation for the sample size was .999 ($p < .001$, agreement rate = 91.1%, $N = 89$). Finally, Cohen's kappa for categorical moderator variables (task type, grade level, and reading level) was .903 ($p < .001$, agreement rate = 93.6%, $N = 421$). Any discrepancies in the ratings were resolved by revisiting the articles and after discussing the coding with the corresponding author.

Meta-Analytic Procedures

The analyses were conducted with the Comprehensive Meta-Analysis program (CMA, Borenstein, Hedges, Higgins, & Rothstein, 2005). The correlations between our predictor variables (PA and MA) and the reading outcomes (reading accuracy, reading fluency, and reading comprehension), as well as information pertinent to task type, grade level, and reading status were coded.

The effect sizes for the studies were displayed by the Pearson's *r* correlation coefficient. A 95% confidence interval (CI) was calculated for each effect size to examine whether the correlation was significantly different from zero. The overall correlation was estimated by calculating a weighted average of the correlations from each study. We used a random-effects model, which rests on the assumption that variation between studies can be systematic and not only due to random error. A sensitivity analysis was also conducted to examine the impact on the overall range of correlations, when studies were removed. Studies were removed one at-a-time to calculate a new overall correlation and the range of this new overall correlation was checked again to make sure the overall correlation was stable (Borenstein et al., 2005). To further examine if the variation in the effect sizes between studies was significant, we performed the Q test of homogeneity (Hedges & Olkin, 2014). A significant value on this test indicates a reliable variability between the correlations in the sample of studies. I^2 was used to determine the magnitude of heterogeneity. I^2 is the proportion of total variation between the effect sizes that is caused by real heterogeneity rather than chance.

For the categorical moderator variables (task type, grade level, and reading ability status), the studies were separated in subsets based on the categories of the moderator variable. The analysis was not conducted when there were fewer than three studies in a category. The degree of differences between the subsets of studies was tested with a Q test and by comparing the correlation magnitude with CIs between the study subsets. Similar to an analysis of variance F test, a Q test would be significant when between-groups difference is statistically larger than within-group difference.

A funnel plot for random-effects models was used to determine the presence of retrieval bias. In the funnel plot, sample size is plotted on the *y*-axis and effect size on the *x*-axis. In the absence of retrieval bias, this plot should be expected to form an inverted funnel. In the presence of bias, the funnel will be asymmetric. To detect retrieval bias, funnel plots are examined for all analyses presented. The trim and fill for random-effects models (Duval & Tweedie, 2000) was used to examine the impact from possible missing studies. The trim-and-fill method imputes values in the funnel plot to make it symmetrical and calculate an estimated overall effect size on this basis.

Results

The literature search and screening process resulted in 64 studies: 32 studies in Chinese and 32 studies in English (see Appendix A, for a list of the studies). Three hundred and 81 separate effect sizes were reported, based on 85 independent samples. In total, 11,138 subjects participated in these studies. The mean age of the participants in Chinese (based on 43 samples that reported the exact age of the participants) was 92.80 months (SD = 23.24 months, range = 52.00–145.72). In turn, the mean age of the participants in English (based on 33 samples that reported the exact age of the participants) was 95.42 months (SD = 19.28 months, range = 57.00–138.30).

Before calculating the average effect size of the correlations between PA, MA, and three reading outcomes, we calculated the correlation between PA and MA, separately for each language. In Chinese, 29 studies and 36 effect sizes described the relationship between PA and MA. The weighted mean correlation was moderate and significant, $r = .34$ (95% CI: .28, .40), z (35) = 10.62, $p < .001$. In turn, 30 studies and 36 effect sizes described the relationship between PA and MA in English. The weighted mean correlation was moderate and significant, $r = .43$ (95% CI: .36, .49), z (35) = 10.98, $p < .001$. The difference across languages was not significant, Q (1) = 3.47, $p = .062$. Because PA correlated significantly with MA in each language, we calculated both the mean effect size of zero-order correlations (see Tables 1–3) as well as the mean effect size of partial correlations (after controlling for each other's effect; see Appendix B).

Mean Effect Size Analyses for Reading Accuracy

PA. Forty-one effect sizes, comprising 5,437 subjects (M sample size = 132.61; SD = 107.95; range = 34–496), described the relationship between PA and reading accuracy in Chinese. The weighted mean correlation was moderate and significant, $r = .30$ (95% CI: .27, .34), z (40) = 16.92, $p < .001$ (see Table 1). The variation in the effect sizes between studies was significant, Q (40) = 63.76, $p = .010$ and $I^2 = 37.27\%$. A sensitivity analysis showed that the overall effect size ranged from .30 (95% CI: .26, .33) to .31 (95% CI: .28, .34). The funnel plot indicated that no studies were missing on either side of the mean. In turn, 41 effect sizes, comprising 5,286 subjects (M sample size = 128.93; SD = 190.06; range = 26–1238), described the relationship between PA and reading accuracy in English. The weighted mean correlation was large and significant, $r = .55$ (95% CI: .50, .59), z (40) = 17.75, $p < .001$ (see Table 1). The variation in the effect sizes between studies was significant, Q (40) = 209.40, $p < .001$ and $I^2 = 80.90\%$. A sensitivity analysis showed that the overall effect size ranged from .54 (95% CI: .49, .58) to .55 (95% CI: .51, .60). The funnel plot indicated that studies were missing on the right side of the mean. In the trim-and-fill analysis (Duval & Tweedie, 2000), six studies were imputed and the adjusted overall mean was .58 (95% CI: .53, .62). A comparison of the correlation coefficients in Chinese and English (see Table 4, top half) revealed that PA had a stronger effect on reading accuracy in English than in Chinese, Q (1) = 58.58, $p < .001$. The difference between languages remained significant, even when partial correlations were considered, $Q(1) = 40.23$, $p < .001$ (see Appendix B, top half).

MA. Forty-one effect sizes, comprising 5,437 subjects (M sample size = 132.61; SD = 107.94; range = 35–496), described the relationship between MA and reading accuracy in Chinese. The weighted mean correlation was moderate and significant, $r = .39$ (95% CI: .36, .43), z (40) = 19.54, $p < .001$ (see Table 1). The variation in the effect sizes between studies was significant, Q (40) = 84.18, $p < .001$ and $I^2 = 52.49\%$. A sensitivity analysis showed the overall effect size ranged from .37 (95% CI: .34, .39) to .38 (95% CI: .36, .41). The funnel plot indicated that no studies were missing on either side of the mean. In turn, 41 effect sizes, comprising 5,286 subjects (M sample size = 128.93; SD = 190.06; range = 26–1238), described the relationship between MA and reading accuracy in English. The weighted mean correlation was moderate and significant, $r = .46$ (95% CI: .40 to .51), z (40) = 14.06, $p < .001$ (see Table 1). The variation in the effect sizes between studies was significant, Q (40) = 223.81, $p < .001$ and $I^2 = 82.13\%$. A sensitivity analysis showed that the overall effect size ranged from .45 (95% CI: .40, .50) to .47 (95% CI: .42, .52). The funnel plot indicated that studies were missing on the right side of the mean. In the trim-and-fill analysis, two studies were imputed and the adjusted overall mean was .48 (95% CI: .43, .53). A comparison of the effects of MA on reading accuracy across the two languages (see Table 4, top half) indicated that MA had a stronger effect on reading accuracy in English than in Chinese, Q (1) = 4.13, $p = .042$. However, the difference failed to reach significance when partial correlations were considered, Q (1) = 1.38, $p = .239$ (see Appendix B, top half).

Mean Effect Size Analyses for Reading Fluency

PA. Six effect sizes, comprising 803 subjects (M sample size = 133.83; SD = 74.69; range = 34–261), described the relationship between PA and reading fluency in Chinese. The weighted mean correlation was small and significant, $r = .26$ (95% CI: .18, .34), z (5) = 6.00, $p < .001$ (see Table 2). The variation in the effect sizes between studies was not significant, Q (5) = 7.30, $p = .200$, and $I^2 = 31.46\%$. A sensitivity analysis showed that the overall effect size ranged from .24 (95% CI: .14, .34) to .30 (95% CI: .23, .37). The funnel plot indicated that studies were missing on the left side of the mean. In the trim-and-fill analysis, a study was imputed and the adjusted overall mean was .26 (95% CI: .18, .33). In turn, 11 effect sizes, comprising 2,462 subjects (M sample size = 223.82; SD = 341.08; range = 43–1238), described the relationship between PA and reading fluency in English. The weighted mean correlation was large and significant, $r = .51$ (95% CI: .44, .58), z (10) = 11.51, $p < .001$ (see Table 2). The variation in the effect sizes between studies was significant, Q (10) = 40.50, $p < .001$

Table 1
Number of Effect Sizes, Effect Size With 95% Confidence Interval, Heterogeneity Statistics, and Moderator Analysis of the Relationship Between Phonological Awareness, Morphological Awareness, and Reading Accuracy in Chinese and English

			Effect size				
Moderator variables	n	k	r	95% CI	z	I^2 (%)	$Q_{between}$
Phonological awareness							58.579***
Chinese	31	41	.302	[.269, .335]	16.915***	37.267*	
Reading status							3.713
Normal		1	.460	[.263, .620]	4.278***	—	
Poor		5	.357	[.240, .463]	5.721***	33.493	
Unselected		35	.293	[.258, .327]	15.820***	36.073*	
Grade level							5.278
Advanced		3	.372	[.285, .454]	7.804***	17.055	
Intermediate		6	.320	[.259, .378]	9.810***	<.001	
Beginning		9	.263	[.211, .314]	9.523***	<.001	
Preschool		18	.317	[.252, .379]	9.124***	56.049**	
Task type							1.953
Complex		21	.256	[.204, .306]	9.398***	28.303	
Simple		31	.307	[.255, .358]	11.008***	61.883***	
English	31	41	.545	[.495, .590]	17.752***	80.898***	
Reading status							1.422
Normal		7	.492	[.407, .568]	9.909***	19.506	
Poor		5	.524	[.378, .645]	6.197***	80.518***	
Unselected		23	.555	[.486, .617]	12.897***	81.036***	
Grade level							2.667
Advanced		6	.495	[.393, .585]	8.355***	78.145***	
Intermediate		8	.470	[.337, .585]	6.278***	63.316***	
Beginning		15	.541	[.455, .617]	10.317***	78.429***	
Preschool		3	.575	[.489, .650]	10.696***	<.001	
Task type							8.576**
Complex		18	.572	[.506, .632]	13.629***	75.509***	
Simple		15	.424	[.343, .498]	9.359***	59.868**	
Morphological awareness							4.138*
Chinese	31	41	.393	[.357, .427]	19.540***	52.485***	
Reading status							8.975*
Normal		1	.640	[.486, .756]	6.522***	—	
Poor		5	.402	[.312, .485]	8.084***	<.001	
Unselected		35	.384	[.346, .420]	18.324***	51.888***	
Grade level							.715
Advanced		3	.436	[.176, .640]	3.163**	87.194***	
Intermediate		6	.413	[.318, .500]	7.826***	54.802	
Beginning		9	.370	[.288, .447]	8.277***	56.487*	
Preschool		18	.379	[.331, .425]	14.210***	28.703	
Task Type I							7.663**
Production		40	.388	[.351, .424]	18.731***	52.938***	
Judgment		19	.288	[.225, .349]	8.565***	63.163***	
Task Type III							4.937
Compounding		35	.376	[.336, .414]	17.164***	43.808**	
Homophone		10	.333	[.217, .439]	5.423***	81.179***	
Homograph		11	.307	[.259, .354]	11.916***	26.867	
English	31	41	.461	[.405, .514]	14.063***	82.128***	
Reading status							1.890
Normal		7	.371	[.211, .511]	4.357***	69.595**	
Poor		5	.467	[.219, .659]	3.501***	91.894***	
Unselected		23	.486	[.405, .559]	10.350***	83.122***	
Grade level							6.269
Advanced		6	.526	[.437, .605]	9.865***	73.393**	
Intermediate		8	.354	[.241, .457]	5.858***	42.859	
Beginning		15	.427	[.323, .520]	7.384***	80.557***	
Preschool		3	.440	[.338, .531]	7.711***	<.001	
Task Type I							5.993*
Production		34	.451	[.387, .511]	12.269***	77.371***	
Judgment		20	.346	[.288, .401]	11.064***	65.968***	
Task Type II							.411
Oral		33	.417	[.350, .481]	11.003***	79.536***	
Written		9	.447	[.383, .506]	12.296***	61.400**	

Note. n = number of studies; k = number of effect sizes; CI = confidence interval; I^2 = the proportion of total variation between the effect size caused by real heterogeneity rather than chance (when <.001 it means that almost all of the observed variance is spurious or there is nothing to explain); $Q_{between}$ = between-groups homogeneity of variance.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2

Number of Effect Sizes, Effect Size With 95% Confidence Interval, Heterogeneity Statistics, and Moderator Analysis of the Relationship Between Phonological Awareness, Morphological Awareness, and Reading Fluency in Chinese and English

			Effect size				
Moderator variables	n	k	r	95% CI	z	I^2 (%)	$Q_{between}$
Phonological awareness							19.511***
Chinese	4	6	.263	[.179, .343]	6.004***	31.457	
Reading status							1.938
Normal		3	.209	[.069, .340]	2.917**	53.362	
Unselected		3	.324	[.231, .411]	6.539***	<.001	
Grade level							1.524
Intermediate		2	.188	[−.044, .401]	1.589	74.524*	
Beginning		1	.250	[.085, .402]	2.934**	–	
Preschool		2	.323	[.226, .414]	6.248***	<.001	
Task type							.248
Complex		5	.232	[.121, .337]	4.050***	30.736	
Simple		6	.266	[.187, .342]	6.377***	32.326	
English	8	11	.509	[.435, .577]	11.512***	75.306***	
Reading status							21.164***
Normal		1	.290	[.079, .476]	2.670**	—	
Poor		3	.373	[.235, .496]	5.043***	45.841	
Unselected		5	.602	[.554, .646]	18.917***	<.001	
Grade level							2.514
Advanced		3	.564	[.437, .668]	7.377***	86.675**	
Intermediate		1	.413	[.233, .566]	4.262***	—	
Beginning		2	.409	[.043, .678]	2.178*	87.084**	
Task type							4.371*
Complex		4	.619	[.570, .663]	18.659***	<.001	
Simple		2	.483	[.344, .601]	6.137***	<.001	
Morphological awareness							.041
Chinese	4	6	.385	[.257, .500]	5.545***	73.556**	
Reading status							1.198
Normal		3	.318	[.139, .478]	3.396**	73.855*	
Unselected		3	.476	[.236, .661]	3.665***	80.623**	
Grade level							2.796
Intermediate		2	.249	[.048, .431]	2.411*	67.134	
Beginning		1	.448	[.302, .574]	5.544***	–	
Preschool		2	.378	[.173, .551]	3.495***	71.341	
Task Type I							.144
Production		12	.350	[.264, .430]	7.567***	71.944***	
Judgment		2	.297	[.019, .533]	2.089*	82.562*	
Task Type III							1.248
Compounding		7	.341	[.226, .446]	5.556***	73.956**	
Homophone		3	.265	[.058, .451]	2.491*	79.327**	
Homograph		4	.398	[.260, .520]	5.309***	67.469*	
English	8	11	.368	[.248, .476]	5.712***	87.650***	
Reading status							9.103*
Normal		1	.030	[−.187, .244]	.268	–	
Poor		3	.286	[−.010, .536]	1.893	86.440**	
Unselected		5	.459	[.279, .608]	4.648***	86.995***	
Grade level							39.065***
Advanced		3	.541	[.462, .612]	11.230***	66.014	
Intermediate		1	.191	[−.009, .376]	1.877	—	
Beginning		2	.074	[−.066, .211]	1.031	<.001	
Task Type I							.909
Production		2	.131	[−.042, .296]	1.484	23.008	
Judgment		10	.243	[.083, .390]	2.955**	88.184***	
Task Type II							3.056
Oral		4	.270	[−.010, .510]	1.890	80.695**	
Written		1	.491	[.447, .532]	18.862***	—	

Note. n = number of studies; k = number of effect sizes; CI = confidence interval; I^2 = the proportion of total variation between the effect size caused by real heterogeneity rather than chance (when <.001 it means that almost all of the observed variance is spurious or there is nothing to explain); $Q_{between}$ = between-groups homogeneity of variance.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Table 3
Number of Effect Sizes, Effect Size With 95% Confidence Interval, Heterogeneity Statistics, and Moderator Analysis of the Relationship Between Phonological Awareness, Morphological Awareness, and Reading Comprehension

			Effect size				
Moderator variables	n	k	r	95% CI	z	I^2 (%)	$Q_{between}$
Phonological awareness							12.300***
Chinese	5	8	.225	[.160, .287]	6.704***	10.151	
Reading status							2.557
Normal		4	.275	[.168, .375]	4.932***	35.196	
Poor		1	.240	[.014, .443]	2.077*	–	
Unselected		3	.163	[.071, .253]	3.439**	<.001	
Grade level							1.349
Advanced		2	.318	[.164, .457]	3.945***	1.775	
Intermediate		3	.213	[.108, .314]	3.914***	9.602	
Beginning		2	.220	[.059, .370]	2.669**	61.461	
Task type							1.278
Complex		6	.258	[.180, .332]	6.338***	5.132	
Simple		6	.179	[.062, .290]	2.997**	62.531*	
English	15	20	.437	[.338, .526]	7.891***	89.569***	
Reading status							.699
Normal		2	.439	[.008, .733]	1.993*	87.407**	
Poor		3	.513	[.362, .638]	5.926***	64.281	
Unselected		11	.433	[.298, .552]	5.799***	86.942***	
Grade level							2.857
Advanced		5	.380	[.230, .513]	4.724***	86.993***	
Intermediate		5	.324	[.070, .539]	2.474*	79.816**	
Beginning		5	.555	[.356, .705]	4.851***	86.056***	
Task type							5.099*
Complex		8	.466	[.328, .585]	5.999***	87.790***	
Simple		4	.248	[.109, .377]	3.452**	<.001	
Morphological awareness							10.337**
Chinese	5	8	.360	[.304, .413]	11.835***	<.001	
Reading status							2.438
Normal		4	.380	[.301, .454]	8.760***	<.001	
Poor		1	.200	[−.028, .408]	1.720	—	
Unselected		3	.362	[.278, .441]	7.925***	<.001	
Grade level							.530
Advanced		2	.360	[.037, .615]	2.173*	77.186*	
Intermediate		3	.337	[.243, .425]	6.680***	<.001	
Beginning		2	.382	[.298, .461]	8.245***	<.001	
Task Type I							1.604
Production		13	.351	[.299, .402]	12.183***	26.189	
Judgment		5	.262	[.128, .388]	3.759***	56.600	
Task Type III							.064
Compounding		5	.324	[.248, .396]	7.924***	<.001	
Homophone		7	.333	[.253, .408]	7.757***	39.212	
Homograph	15	5	.343	[.198, .473]	4.475***	70.663**	
English		20	.534	[.444, .613]	9.893***	89.849***	
Reading status							.855
Normal		2	.432	[.142, .654]	2.841**	73.414	
Poor		3	.587	[.205, .814]	2.834**	94.197***	
Unselected		11	.552	[.420, .661]	7.039***	89.364***	
Grade level							3.231
Advanced		5	.536	[.418, .637]	7.639***	84.694***	
Intermediate		5	.386	[.250, .508]	5.264***	38.331	
Beginning		5	.522	[.266, .709]	3.707***	90.578***	
Task Type I							.365
Production		11	.490	[.334, .620]	5.563***	88.451***	
Judgment		14	.443	[.415, .471]	27.154***	2.395	
Task Type II							1.055
Oral		7	.530	[.358, .668]	5.362***	91.048***	
Written		8	.443	[.411, .474]	23.772***	4.465	

Note. n = number of studies; k = number of effect sizes; CI = confidence interval; I^2 = the proportion of total variation between the effect size caused by real heterogeneity rather than chance (when <.001 it means that almost all of the observed variance is spurious or there is nothing to explain); $Q_{between}$ = between-groups homogeneity of variance.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4
Language Comparisons Within Phonological Awareness and Morphological Awareness and Meta-Linguistic Awareness Comparisons Within English and Chinese

	r-values						
Awareness	CH	EN	PA	MA	*r*-values defferences	$Q_{between}$	*p*
Phonological awareness							
Reading accuracy	.302	.545			EN > CH	58.579	<.001
Reading fluency	.263	.509			EN > CH	19.511	<.001
Reading Comprehension	.225	.437			EN > CH	12.300	<.001
Morphological awareness							
Reading accuracy	.393	.461			EN > CH	4.138	.042
Reading fluency	.385	.368			n.s.	.041	.840
Reading Comprehension	.360	.534			EN > CH	10.337	.001
Chinese							
Reading accuracy			.302	.393	PA < MA	13.396	<.001
Reading fluency			.263	.385	n.s.	2.533	.111
Reading Comprehension			.225	.360	PA < MA	10.059	.002
English							
Reading accuracy			.545	.461	PA > MA	5.107	.024
Reading fluency			.509	.368	PA > MA	4.478	.034
Reading Comprehension			.437	.534	n.s.	2.238	.135

Note. CH = Chinese; EN = English; PA = phonological awareness; MA = morphological awareness.

and I^2 = 75.31%. A sensitivity analysis showed that the overall effect size ranged from .49 (95% CI: .42, .56) to .53 (95% CI: .46, .59). The funnel plot indicated that no studies were missing on either side of the mean. A comparison of the correlation coefficients across the two languages (see Table 4, top half) revealed that PA had a stronger effect on reading fluency in English than in Chinese, Q (1) = 19.51, $p < .001$. A similar finding was obtained with partial correlations, Q (1) = 22.74, $p < .001$ (see Appendix B, top half).

MA. Six effect sizes, comprising 803 subjects (*M* sample size = 133.83; *SD* = 74.53; range = 35–261), described the relationship between MA and reading fluency in Chinese. The weighted mean correlation was moderate and significant, $r = .39$ (95% CI: .26, .50), z (5) = 5.55, $p < .001$ (see Table 2). The variation in the effect sizes between studies was significant, Q (5) = 18.91, $p = .002$ and I^2 = 73.56%. A sensitivity analysis showed that the overall effect size ranged from .34 (95% CI: .23, .44) to .43 (95% CI: .30, .53). The funnel plot indicated that studies were missing on the left side of the mean. In the trim-and-fill analysis, a study was imputed and the adjusted overall mean was .34 (95% CI: .20, .47). In turn, 11 effect sizes, comprising 2,462 subjects (*M* sample size = 223.82; *SD* = 341.08; range = 43–1238), described the relationship between MA and reading fluency in English. The weighted mean correlation was moderate and significant, $r = .37$ (95% CI: .25, .48), z (10) = 5.71, $p < .001$ (see Table 2). The variation in the effect sizes between studies was significant, Q (10) = 80.97, $p < .001$ and I^2 = 87.65%. A sensitivity analysis showed that the overall effect size ranged from .34 (95% CI: .21, .45) to .40 (95% CI: .28, .50). The funnel plot indicated that studies were missing on the right side of the mean. In the trim-and-fill analysis, two studies were imputed and the adjusted overall mean was .43 (95% CI: .32, .53). A comparison of the effects of MA on reading fluency across the two languages (see Table 4, top half) indicated no significant differences, Q (1) = 0.04, $p = .840$. A similar finding was obtained with partial correlations, Q (1) = 0.30, $p = .583$ (see Appendix B, top half).

Mean Effect Size Analyses for Reading Comprehension

PA. Eight effect sizes, comprising 1,013 subjects (*M* sample size = 126.63; *SD* = 73.07; range = 64–290), described the relationship between PA and reading comprehension in Chinese. The weighted mean correlation was small and significant $r = .23$ (95% CI: .16, .29), z (7) = 6.70, $p < .001$ (see Table 3). The variation in the effect sizes between studies was not significant, Q (7) = 7.80, $p = .351$ and I^2 = 10.15%. A sensitivity analysis showed that the overall effect size was in the range of .21 (95% CI: .14, .27) to .25 (95% CI: .18, .32). The funnel plot indicated that studies were missing on the left side of the mean. In the trim-and-fill analysis, a study was imputed and the adjusted overall mean was .21 (95% CI: .14, .28). In turn, 20 effect sizes, comprising 3,419 subjects (*M* sample size = 170.95; *SD* = 262.90; range = 26–1238), described the relationship between PA and reading comprehension in English. The weighted mean correlation was moderate and significant, $r = .44$ (95% CI: .34, .53), z (19) = 7.89, $p < .001$ (see Table 3). The variation in the effect sizes between studies was significant, Q (19) = 182.14, $p < .001$ and I^2 = 89.57%. A sensitivity analysis showed that the overall effect size ranged from .42 (95% CI: .32, .51) to .45 (95% CI: .35, .54). The funnel plot indicated that studies were missing on the right side of the mean. In the trim-and-fill analysis, three studies were imputed and the adjusted overall mean was .48 (95% CI: .39, .56). A comparison of the correlation coefficients across the two languages (see Table 4, top half) revealed that PA had a stronger effect on reading comprehension in English than in Chinese, Q (1) = 12.30, $p < .001$) However, the difference failed to reach significance with partial correlations, Q (1) = 2.17, $p = .140$ (see Appendix B, top half).

MA. Eight effect sizes, comprising 1,013 subjects (*M* sample size = 126.63; *SD* = 73.07; range = 64–290), described the relationship between MA and reading comprehension in Chinese. The weighted mean correlation was moderate and significant, $r =$

.36 (95% CI: .30, .41), z (7) = 11.84, $p < .001$ (see Table 3). The variation in the effect sizes between studies was not significant, Q (7) = 5.59, p = .588 and $I^2 < .001\%$. A sensitivity analysis showed that the overall effect size ranged from .35 (95% CI: .28, .41) to .37 (95% CI: .31, .43). The funnel plot indicated that studies were missing on the right side of the mean. In the trim-and-fill analysis, two studies were imputed and the adjusted overall mean was .38 (95% CI: .33, .43). In turn, 20 effect sizes, comprising 3,419 subjects (M sample size = 170.95; SD = 262.90; range = 26–1238), described the relationship between MA and reading comprehension in English. The weighted mean correlation was large and significant, r = .53 (95% CI: .44, .61), z (19) = 9.89, $p < .001$ (see Table 3). The variation in the effect sizes between studies was significant, Q (19) = 187.17, $p < .001$ and I^2 = 89.85%. A sensitivity analysis showed that the overall effect size ranged from .51 (95% CI: .42, .59) to .55 (95% CI: .47, .63). The funnel plot indicated that studies were missing on the right side of the mean. In the trim-and-fill analysis, two studies were imputed, and the adjusted overall mean was .57 (95% CI: .48, .64). A comparison of the correlation coefficients across the two languages (see Table 4, top half) revealed that MA had a stronger effect on reading comprehension in English than in Chinese, Q (1) = 10.34, p = .001. However, the difference failed to reach significance with partial correlations, Q (1) = 2.18, p = .139 (see Appendix B, top half).

Comparing the Effects of PA and MA Across Reading Outcomes Within Each Language

To examine whether PA correlated more strongly with reading accuracy, fluency, and comprehension than MA (or vice versa), we performed a Q-test, separately for each language (see Table 4, bottom half). In Chinese, MA correlated more strongly with reading accuracy, Q (1) = 13.40, $p < .001$, and reading comprehension, Q (1) = 10.06, p = .002, than PA. No significant difference was observed for reading fluency, Q (1) = 2.53, p = .11). Similar findings were obtained with partial correlations (see Appendix B, bottom half). In contrast, in English, PA correlated more strongly with reading accuracy, Q (1) = 5.11, p = .024, and reading fluency, Q (1) = 4.48, p = .034, than MA. No significant difference was observed for reading comprehension, Q (1) = 2.24, p = .135. When we repeated the analyses with partial correlations, MA correlated more strongly with reading comprehension in English than PA, Q (1) = 6.01, p = .014 (see Appendix B, bottom half).

Moderator Analyses

The results of the moderator analyses are shown in Tables 1, 2, and 3.

Task type. For PA, only the difference in the relationship of simple and complex PA tasks with reading in English was significant. Complex PA tasks correlated more strongly with reading accuracy, Q (1) = 8.58, p = .003, reading fluency, Q (1) = 4.37, p = .037, and reading comprehension, Q (1) = 5.10, p = .024, than simple PA tasks. For MA, the only significant difference was found when comparing production and judgment tasks. The production tasks correlated more strongly with reading accuracy than the judgment tasks in both Chinese, Q (1) = 7.66, p = .006, and English, Q (1) = 5.99, p = .014.

Grade level. A significant difference was observed only in the relationship between MA and reading fluency in English, Q (2) = 39.07, $p < .001$. The correlation was significant among advanced readers (r = .54), but not among beginning (r = .07) or intermediate (r = .19) readers.

Reading ability status. In English, statistically significant differences between groups of readers were observed when reading fluency was the reading outcome, in the correlations with both PA, Q (2) = 21.16, $p < .001$, and MA, Q (2) = 9.10, p = .011. Studies with unselected samples produced the strongest correlations (PA: r = .60; MA: r = .46), followed by studies with poor readers (PA: r = .37; MA: r = .29), and finally by studies with normal readers (PA: r = .29; MA: r = .03). In Chinese, a significant difference was found in the relationship between MA and reading accuracy, Q (2) = 8.98, p = .011, with the strongest correlations obtained in studies with normal readers (r = .64), followed by studies with poor readers (r = .40), and finally by studies with unselected samples of readers (r = .38).

Discussion

The primary goal of this meta-analysis was to examine if the size of the relationship of PA and MA with different reading outcomes varies between English and Chinese. Given the linguistic features of each language, we hypothesized that a stronger relationship between PA and reading would be observed in English than in Chinese, and a stronger relationship between MA and reading would be observed in Chinese than in English. In line with our expectation and with the findings of previous studies (e.g., McBride-Chang, Bialystok, Chong, & Li, 2004; McBride-Chang, Cho et al., 2005, 2013; Tong & McBride-Chang, 2010), PA was more strongly related to all reading outcomes in English than in Chinese (however, the difference in reading comprehension disappeared when partial correlations were considered). This reinforces the argument put forward by several researchers that PA is fundamental to word reading in English (e.g., Bowey, 2005; Scarborough, 1998). In contrast, the low percentage of characters with a regular phonetic radical in Chinese (i.e., 23–26% when the tone is taken into account; Chung & Leung, 2008), renders the use of the phonetic radical to character reading inefficient, thus possibly downgrading the importance of PA in Chinese.

However, PA still produced significant correlations with reading in Chinese (the average correlations with reading accuracy and fluency are similar to those reported in Song et al.'s meta-analysis). A significant correlation would be expected based on the fact that 80% of Chinese characters contain a phonetic radical that provides some clues to character's pronunciation. Perhaps the percentage of these characters whose pronunciation is consistent with the phonetic radical (23–26%; Chung & Leung, 2008), albeit relatively low, is sufficient enough to produce a significant correlation. This is in line with the findings of some studies showing that the knowledge of phonetic radicals assists Chinese children in learning to read (e.g., Ho & Bryant, 1997; Wu, Zhou, & Shu, 1999).

A significant correlation would also be expected if we take into account how Chinese children in mainland China and Taiwan learn to read. Specifically, in mainland China, children are introduced to a phonetic alphabet called Pinyin to assist them in learning new characters. The Pinyin system borrows English letters to represent individual phonemes. In turn, in Taiwan, children are presented with a

phonetic alphabet called Zhuyin Fuhao. Zhuyin Fuhao roughly transcribes spoken sounds at the onset-rime level and is printed alongside the new characters in the children's textbooks. Although children in Hong Kong are not exposed to a phonetic alphabet, they learn English from a very young age (at least those from relatively affluent families). Thus, exposure to and practice with PA tasks happens indirectly through learning to read English words.

Notice also that the correlations between PA and Chinese reading remained significant (albeit weak), after partialing out the effects of MA (see Appendix B). This suggests that PA plays an independent role in learning to read Chinese that is not completely overlapping with that of MA. This is important because the nonsignificant effects of PA on Chinese reading in previous studies (e.g., McBride-Chang, Cho et al., 2005; Yeung et al., 2011) were attributed to the inclusion of MA in the same models.

In contrast to our expectation, MA correlated more strongly with reading accuracy and comprehension in English than in Chinese (although none of these differences remained significant in the analyses with partial correlations). In addition, with one exception (the relationship of MA with reading fluency in English, which was stronger among advanced readers than among beginning or intermediate readers), grade level did not moderate the relationship between MA and reading. Taken together, these findings suggest that MA is equally important for reading in English (or even more important than Chinese when zero-order correlations are taken into account) and that the relationships can be found even among young children. This is in line with Nunes and Hatano's (2004) conclusion that irrespective of differences between writing systems, MA is important for learning to read.

However, in Chinese, MA correlated more strongly with reading accuracy and comprehension than PA. This can be attributed to the richness of homophones in Chinese (Kuo & Anderson, 2006). It has been estimated that a spoken Mandarin syllable represents an average of five morphemes (Packard, 2000), whereas a spoken Cantonese syllable represents an average of three morphemes (Chow et al., 2008). Given the one-to-many relationship between a syllable and a morpheme in Mandarin and Cantonese, it is not always reliable to distinguish words with the same pronunciation by simply relying on PA. In addition, because the way morphemes are combined to form words in Chinese tends to be regular and informative (i.e., the meaning of most Chinese compound words is predictable from the meaning of their constituent morphemes), Chinese children rely on their MA skills to recognize words (e.g., Li et al., 2002; Liu & McBride-Chang, 2010; McBride-Chang, Cho et al., 2005; Shu et al., 2006; Tong et al., 2011; Wei et al., 2014). In contrast, in English, PA correlated more strongly with reading accuracy and fluency than did MA. This suggests that although English-speaking children may pay attention to the internal structure of words, they rely more heavily on PA to decode words and to read fluently. An alternative explanation may relate to the nature of the reading tasks.[4] Specifically, because most word reading tasks involve reading words in isolation, without meaningful context, this may have inflated the role of PA in word reading. In addition, in many English word reading tasks used in the studies incorporated in our meta-analysis, particularly those targeting younger children, the stimuli were one morpheme words. This may have also lessened the relationship of MA with word reading. Finally, although it is tempting to argue that the way reading is taught in English may have given an unfair advantage to PA (given that most teachers in North America teach PA explicitly; learning of morphology is achieved implicitly), the fact that MA in Chinese was a stronger correlate of reading accuracy and reading comprehension in the absence of any explicit teaching of morphology by Chinese teachers suggests that instructional practices are not likely the key factor determining the size of the relationship between these linguistic skills and reading.

From a theoretical point of view, our findings suggest that learning to read across different writing systems involves the same set of mappings between orthographic (written) and phonological (spoken)/semantic forms of words (e.g., Seidenberg, 2011). However, the "division of labor" between these processes differs by writing system. Both PA and MA correlated significantly with the reading outcomes in both languages. However, PA plays a stronger role than MA in word reading in English because the spelling-to-sound mappings are relatively systematic (or at least not as ambiguous as in Chinese). In Chinese, MA is a stronger correlate of reading accuracy and comprehension than PA, because the basic graphic unit, the character, represents a morpheme (not a phoneme).

Some limitations of the present study are worth mentioning. First, we ran our analyses using studies conducted only in Chinese and English. This was done not only because English and Chinese represent different writing systems, but also because they differ on important linguistic characteristics that have direct implications for the role of PA or MA in reading. However, we acknowledge that English is an atypical alphabetic orthography and our results may not generalize to other alphabetic languages with a more transparent orthography. Second, we considered studies in our meta-analysis that assessed both PA and MA in the same study. Although the issue of including studies examining both skills or studies examining either skill in a meta-analysis is still a matter of debate (e.g., Kulinskaya, Morgenthaler, & Staudte, 2008), we made this decision to gain a better control over the possible effects of confounding variables (e.g., sample characteristics) in the size of the correlations and to be able to compare our findings to those of previous meta-analyses that used a similar approach (e.g., Song et al., 2016; Swanson et al., 2003). However, we acknowledge that this has reduced the number of studies that were considered in the meta-analysis. Third, we did not examine the role of dialect (Mandarin vs. Cantonese) or script (simplified vs. traditional) in the relationship between PA, MA, and reading in Chinese. Had we examined their role, we would not be able to run some of the analyses because of the small number of studies conducted in either dialect or script. Fourth, we did not include any control variables (e.g., socioeconomic status, IQ, vocabulary) in our study. This was done for a methodological reason since there could be several potential control variables and not all studies incorporated in our meta-analysis assessed the same control variables. Fifth, information on how the children were taught to read was missing from most studies and for this reason we refrained from making any generalizations about reading instruction. Whether the observed differences in the size of the correlations between PA, MA and reading in English reflect the way children are taught how to read is a question that remains to be answered in a future study. Sixth, although grade level and reading ability status may be somewhat confounded, we could not test the interaction between the two because we had a very small number of effect sizes in the poor and normal readers' groups. Finally, we

[4] We thank the anonymous reviewer for this explanation.

acknowledge that our classification of the MA tasks is not the only one. Deacon et al. (2008), for example, proposed a taxonomy of MA tasks that takes into account not only the format of presentation (oral vs. written), but also the content (e.g., whether phonological or orthographic shifts happen when morphemes are added to a base or stem) and process (e.g., whether the task requires explicit or only implicit knowledge).

Psychoeducational Implications

The findings of this meta-analysis have some important psychoeducational implications. First, given that both PA and MA were found to be significant correlates of reading in both languages, tasks of both skills may be used to screen for reading difficulties. Currently, most screening batteries in English and Chinese include measures of PA, but not measures of MA. Second, given that PA and MA correlated significantly with each other in both languages, instruction in either skill may facilitate the learning of the other. This further suggests that some children with weak PA may be taught MA to compensate (e.g., Deacon et al., 2008).

Conclusion

To conclude, our meta-analysis is the first one to document the differential relationship of PA and MA with different reading outcomes across two writing systems (alphabetic and logographic). The results suggest that significant differences across the two languages included in this meta-analysis could only be detected in the role of PA (based on the results with partial correlations). Specifically, PA was a stronger correlate of reading accuracy and fluency in English than in Chinese. However, when we look at each language separately, variation is rather predictable. Because in Chinese the mapping from spelling to sound is syllable-based with no constituent parts of a character corresponding to phonemes, MA plays a more important role in reading than PA. In contrast, because in English letters correspond to sounds, PA plays a more important role in word reading than MA.

References

*An asterisk is put in front of studies that have been used in the meta-analysis.

*Apel, K., & Lawrence, J. (2011). Contributions of morphological awareness skills to word-level reading and spelling in first-grade children with and without speech sound disorder. *Journal of Speech, Language, and Hearing Research, 54,* 1312–1327. http://dx.doi.org/10.1044/1092-4388(2011/10-0115)

*Apel, K., & Thomas-Tate, S. (2009). Morphological awareness skills of fourth-grade African American students. *Language, Speech, and Hearing Services in Schools, 40,* 312–324. http://dx.doi.org/10.1044/0161-1461(2009/08-0015)

*Apel, K., Wilson-Fowler, E. B., Brimo, D., & Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. *Reading and Writing, 25,* 1283–1305. http://dx.doi.org/10.1007/s11145-011-9317-8

*Berninger, V. W., Abbott, R. D., Thomson, J., Wagner, R., Swanson, H. L., Wijsman, E. M., & Raskind, W. (2006). Modeling phonological core deficits within a working memory architecture in children and adults with developmental dyslexia. *Scientific Studies of Reading, 10,* 165–198. http://dx.doi.org/10.1207/s1532799xssr1002_3

*Berninger, V. W., Raskind, W., Richards, T., Abbott, R., & Stock, P. (2008). A multidisciplinary approach to understanding developmental dyslexia within working-memory architecture: Genotypes, phenotypes, brain, and instruction. *Developmental Neuropsychology, 33,* 707–744. http://dx.doi.org/10.1080/87565640802418662

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive meta-analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research, 80,* 144–179. http://dx.doi.org/10.3102/0034654309359353

Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 155–172). Oxford, UK: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch9

Carlisle, J. F. (1995). Morphological awareness and early reading achievement. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 189–209). Hillsdale, NJ: Erlbaum.

Carlisle, J. F., McBride-Chang, C., Nagy, W., & Nunes, T. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly, 45,* 464–487. http://dx.doi.org/10.1598/RRQ.45.4.5

*Carlisle, J. F., & Nomanbhoy, D. M. (1993). Phonological and morphological awareness in first graders. *Applied Psycholinguistics, 14,* 177–195. http://dx.doi.org/10.1017/S0142716400009541

Chao, Y. R. (1976). *Aspects of Chinese sociolinguistics: Essays* (Vol. 9). Palo Alto, CA: Stanford University Press.

*Chen, X., Hao, M. L., Geva, E., Zhu, J., & Shu, H. (2009). The role of compound awareness in Chinese children's vocabulary acquisition and character reading. *Reading and Writing, 22,* 615–631. http://dx.doi.org/10.1007/s11145-008-9127-9

*Cheung, H., Wong, S. W. L., Penney, T. B., Chung, K. K. H., McBride-Chang, C., & Ho, C. S. H. (2010). Speech perception, metalinguistic awareness, reading, and vocabulary in Chinese-English bilingual children. *Journal of Educational Psychology, 102,* 367–380. http://dx.doi.org/10.1037/a0017850

*Chik, P. P. M., Ho, C. S. H., Yeung, P. S., Chan, D. W. O., Chung, K. K. H., Luan, H., . . . Lau, W. S. Y. (2012). Syntactic skills in sentence reading comprehension among Chinese elementary school children. *Reading and Writing, 25,* 679–699. http://dx.doi.org/10.1007/s11145-010-9293-4

Cho, J.-R., Chiu, M. M., & McBride-Chang, C. (2011). Morphological awareness, phonological awareness, and literacy development in Korean and English: A 2-year longitudinal study. *Scientific Studies of Reading, 15,* 383–408. http://dx.doi.org/10.1080/10888438.2010.487143

Chomsky, C. (1970). Reading, writing, and phonology. *Harvard Educational Review, 40,* 287–309. http://dx.doi.org/10.17763/haer.40.2.y7u0242x76w05624

*Chow, B. W. Y., Ho, C. S. H., Wong, S. W. L., Waye, M. M. Y., & Bishop, D. V. M. (2011). Genetic and environmental influences on Chinese language and reading abilities. *PLoS ONE, 6,* e16640. http://dx.doi.org/10.1371/journal.pone.0016640

Chow, B. W., McBride-Chang, C., Cheung, H., & Chow, C. S. (2008). Dialogic reading and morphology training in Chinese children: Effects on language and literacy. *Developmental Psychology, 44,* 233–244. http://dx.doi.org/10.1037/0012-1649.44.1.233

Chung, F. H., & Leung, M. T. (2008). Data analysis of Chinese characters in primary school corpora of Hong Kong and mainland China: Preliminary theoretical interpretations. *Clinical Linguistics & Phonetics, 22,* 379–389. http://dx.doi.org/10.1080/02699200701776757

*Chung, K. K., & Ho, C. S. (2010). Second language learning difficulties in Chinese children with dyslexia: What are the reading-related cognitive skills that contribute to English and Chinese word reading? *Journal*

of Learning Disabilities, 43, 195–211. http://dx.doi.org/10.1177/0022219409345018

*Chung, K. K. H., & McBride-Chang, C. (2011). Executive functioning skills uniquely predict Chinese word reading. *Journal of Educational Psychology, 103,* 909–921. http://dx.doi.org/10.1037/a0024744

*Chung, K. K. H., McBride-Chang, C., Wong, S. W., Cheung, H., Penney, T. B., & Ho, C. S. H. (2008). The role of visual and auditory temporal processing for Chinese children with developmental dyslexia. *Annals of Dyslexia, 58,* 15–35. http://dx.doi.org/10.1007/s11881-008-0015-4

*Cunningham, A. J., & Carroll, J. M. (2015). Early predictors of phonological and morphological awareness and the link with reading: Evidence from children with different patterns of early deficit. *Applied Psycholinguistics, 36,* 509–531. http://dx.doi.org/10.1017/S0142716413000295

*Deacon, S. H. (2012). Sounds, letters and meanings: The independent influences of phonological, morphological and orthographic skills on early word reading accuracy. *Journal of Research in Reading, 35,* 456–475. http://dx.doi.org/10.1111/j.1467-9817.2010.01496.x

*Deacon, S. H., Benere, J., & Pasquarella, A. (2013). Reciprocal relationship: Children's morphological awareness and their reading accuracy across grades 2 to 3. *Developmental Psychology, 49,* 1113–1126. http://dx.doi.org/10.1037/a0029474

*Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics, 25,* 223–238. http://dx.doi.org/10.1017/S0142716404001110

Deacon, S. H., Parrila, R., & Kirby, J. R. (2008). A review of the evidence on morphological processing in dyslexics and poor readers: A strength or weakness? In F. Manis, A. Fawcett, G. Reid, & L. Siegel (Eds.), *Sage handbook of dyslexia* (pp. 212–238). London, UK: Sage. http://dx.doi.org/10.4135/9780857020987.n11

*Deacon, S. H., Wade-Woolley, L., & Kirby, J. (2007). Crossover: The role of morphological awareness in French immersion children's reading. *Developmental Psychology, 43,* 732–746. http://dx.doi.org/10.1037/0012-1649.43.3.732

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167–188. http://dx.doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250–287. http://dx.doi.org/10.1598/RRQ.36.3.2

*Farran, L. K., Bingham, G. E., & Matthews, M. W. (2012). The relationship between language and reading in bilingual English-Arabic children. *Reading and Writing, 25,* 2153–2181. http://dx.doi.org/10.1007/s11145-011-9352-5

*Fraser, J., & Conti-Ramsden, G. (2008). Contribution of phonological and broader language skills to literacy. *International Journal of Language & Communication Disorders, 43,* 552–569. http://dx.doi.org/10.1080/13682820701778069

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia, 60,* 183–208. http://dx.doi.org/10.1007/s11881-010-0041-x

Grigorakis, I. (2014). *The development of children's metamorphological abilities from Kindergarten to Grade 2 and its relationship with reading and spelling acquisition* (Unpublished doctoral dissertation). University of Crete, Rethymno, Greece.

Hanley, J. R. (2005). Learning to read in Chinese. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 316–335). Oxford, UK: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch17

Hedges, L. V., & Olkin, I. (2014). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

*Ho, C. S.-H. (2014). Preschool predictors of dyslexia status in Chinese first graders with high or low familial risk. *Reading and Writing, 27,* 1673–1701. http://dx.doi.org/10.1007/s11145-014-9515-2

Ho, C. S.-H., & Bryant, P. (1997). Phonological skills are important in learning to read Chinese. *Developmental Psychology, 33,* 946–951. http://dx.doi.org/10.1037/0012-1649.33.6.946

*Ho, C. S.-H., Leung, M. T., & Cheung, H. (2011). Early difficulties of Chinese preschoolers at familial risk for dyslexia: Deficits in oral language, phonological processing skills, and print-related skills. *Dyslexia: An International Journal of Research and Practice, 17,* 143–164. http://dx.doi.org/10.1002/dys.429

*Hu, C. F. (2013). Predictors of reading in children with Chinese as a first language: A developmental and cross-linguistic perspective. *Reading and Writing, 26,* 163–187. http://dx.doi.org/10.1007/s11145-012-9360-0

Institute of Language Teaching and Research of China. (Ed.) (1986). *A frequency dictionary of modern Chinese.* Beijing, China: Beijing Language Institute Press.

*Jarmulowicz, L., Hay, S. E., Taran, V. L., & Ethington, C. A. (2008). Fitting derivational morphophonology into a developmental model of reading. *Reading and Writing, 21,* 275–297. http://dx.doi.org/10.1007/s11145-007-9073-y

*Jarmulowicz, L., Taran, V. L., & Hay, S. E. (2007). Third graders' metalinguistic skills, reading skills, and stress production in derived English words. *Journal of Speech, Language, and Hearing Research, 50,* 1593–1605. http://dx.doi.org/10.1044/1092-4388(2007/107)

*Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology, 77,* 30–60. http://dx.doi.org/10.1006/jecp.1999.2553

Kamhi, A., & Catts, H. W. (2012). *Language and reading disabilities* (3rd ed.). Boston, MA: Pearson.

Katzir, T., Kim, Y.-S., Wolf, M., Kennedy, B., Lovett, M., & Morris, R. (2006). The relationship of spelling recognition, RAN, and phonological awareness to reading skills in older poor readers and younger reading-matched controls. *Reading and Writing, 19,* 845–872. http://dx.doi.org/10.1007/s11145-006-9013-2

Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101,* 765–778. http://dx.doi.org/10.1037/a0015956

*Kim, Y.-S., Apel, K., & Al Otaiba, S. (2013). The relation of linguistic awareness and vocabulary to word reading and spelling for first-grade students participating in response to intervention. *Language, Speech, and Hearing Services in Schools, 44,* 337–347. http://dx.doi.org/10.1044/0161-1461(2013/12-0013)

*Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading and Writing, 25,* 389–410. http://dx.doi.org/10.1007/s11145-010-9276-5

*Kruk, R. S., & Bergman, K. (2013). The reciprocal relations between morphological processes and reading. *Journal of Experimental Child Psychology, 114,* 10–34. http://dx.doi.org/10.1016/j.jecp.2012.09.014

Kulinskaya, E., Morgenthaler, S., & Staudte, R. G. (2008). *Meta analysis: A guide to calibrating and combining statistical evidence* (Vol. 756). West Sussex, England: Wiley.

Kuo, K.-J., & Anderson, R. C. (2006). Morpholgical awareness and learning to read: A cross-language perspective. *Educational Psychologist, 41,* 161–180. http://dx.doi.org/10.1207/s15326985ep4103_3

*Lam, F. W. F., McBride-Chang, C., Lam, C. C. C., Wong, S. W. L., Chow, Y., & Doo, S. (2008). Towards early identification of dyslexia in Chinese preschool children: A study on reading and cognitive profile in children with genetic risk of dyslexia in Hong Kong. *Hong Kong Journal of Paediatrics, 13,* 90–98.

*Lei, L., Pan, J., Liu, H., McBride-Chang, C., Li, H., Zhang, Y., . . . Shu, H. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *Journal of Child Psychology and Psychiatry, 52,* 212–220. http://dx.doi.org/10.1111/j.1469-7610.2010.02311.x

Li, H., Dronjic, V., Chen, X. I., Li, Y., Cheng, Y., & Wu, X. (in press). Morphological awareness as a function of semantics, phonology, and orthography and as a predictor of reading comprehension in Chinese. *Journal of Child Language.* http://dx.doi.org/10.1017/S0305000916000477

*Li, H., Shu, H., McBride-Chang, C., Liu, H. Y., & Peng, H. (2012). Chinese children's character recognition: Visuo-orthographic, phonological processing and morphological skills. *Journal of Research in Reading, 35,* 287–307. http://dx.doi.org/10.1111/j.1467-9817.2010.01460.x

*Li, L., & Wu, X. (2015). Effects of metalinguistic awareness on reading comprehension and the mediator role of reading fluency from grades 2 to 4. *PLoS ONE, 10,* e0114417. http://dx.doi.org/10.1371/journal.pone.0114417

Li, W., Anderson, R. C., Nagy, W., & Zhang, H. (2002). Facets of metalinguistic awareness that contribute to Chinese literacy. In W. Li, J. S. Gaffney, & J. L. Packard (Eds.), *Chinese children's reading acquisition: Theoretical and pedagogical issues* (pp. 87–106). Boston, MA: Kluwer Academic.

*Lin, D., McBride-Chang, C., Aram, D., & Levin, I. (2011). Mother-child joint writing in Chinese kindergarteners: Metalinguistic awareness, maternal mediation and literacy acquisition. *Journal of Research in Reading, 34,* 426–442. http://dx.doi.org/10.1111/j.1467-9817.2010.01446.x

*Lin, D., McBride-Chang, C., Aram, D., Shu, H., Levin, I., & Cho, J. R. (2012). Maternal mediation of word writing in Chinese across Hong Kong and Beijing. *Journal of Educational Psychology, 104,* 121–137. http://dx.doi.org/10.1037/a0025383

*Liu, D., Chen, X., & Chung, K. K. H. (2015). Performance in a visual search task uniquely predicts reading abilities in third-grade Hong Kong Chinese children. *Scientific Studies of Reading, 19,* 307–324. http://dx.doi.org/10.1080/10888438.2015.1030749

*Liu, D., Chung, K. K. H., Zhang, Y. M., & Lu, Z. (2014). Sensitivity to the positional information of morphemes inside Chinese compound words and its relationship with word reading. *Reading and Writing, 27,* 431–450. http://dx.doi.org/10.1007/s11145-013-9451-6

*Liu, D., & McBride-Chang, C. (2014). Morphological structure processing during word recognition and its relationship to character reading among third-grade Chinese children. *Journal of Psycholinguistic Research, 43,* 715–735. http://dx.doi.org/10.1007/s10936-013-9275-1

*Liu, P. D., & McBride-Chang, C. (2010). What is morphological awareness? Tapping lexical compounding awareness in Chinese third graders. *Journal of Educational Psychology, 102,* 62–73. http://dx.doi.org/10.1037/a0016933

*Liu, P. D., McBride-Chang, C., Wong, T. T. Y., Shu, H., & Wong, A. M. Y. (2013). Morphological awareness in Chinese: Unique associations of homophone awareness and lexical compounding to word reading and vocabulary knowledge in Chinese children. *Applied Psycholinguistics, 34,* 755–775. http://dx.doi.org/10.1017/S014271641200001X

*Mahony, D., Singson, M., & Mann, V. (2000). Reading ability and sensitivity to morphological relations. *Reading and Writing, 12,* 191–218. http://dx.doi.org/10.1023/A:1008136012492

McBride-Chang, C., Bialystok, E., Chong, K. K., & Li, Y. (2004). Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology, 89,* 93–111. http://dx.doi.org/10.1016/j.jecp.2004.05.001

*McBride-Chang, C., Cheung, H., Chow, B. W. Y., Chow, C. S. L., & Choi, L. (2006). Metalinguistic skills and vocabulary knowledge in Chinese (L1) and English (L2). *Reading and Writing, 19,* 695–716. http://dx.doi.org/10.1007/s11145-005-5742-x

*McBride-Chang, C., Cho, J. R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., . . . Muse, A. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology, 92,* 140–160. http://dx.doi.org/10.1016/j.jecp.2005.03.009

McBride-Chang, C., Chung, K. K., & Tong, X. (2011). Copying skills in relation to word reading and writing in Chinese children with and without dyslexia. *Journal of Experimental Child Psychology, 110,* 422–433. http://dx.doi.org/10.1016/j.jecp.2011.04.014

McBride-Chang, C., & Ho, C. S.-H. (2000). Naming speed and phonological awareness in Chinese children: Relations to reading skills. *Journal of Psychology in Chinese Societies, 1,* 93–108.

*McBride-Chang, C., Lam, F., Lam, C., Chan, B., Fong, C. Y. C., Wong, T. T. Y., & Wong, S. W. L. (2011). Early predictors of dyslexia in Chinese children: Familial history of dyslexia, language delay, and cognitive profiles. *Journal of Child Psychology and Psychiatry, 52,* 204–211. http://dx.doi.org/10.1111/j.1469-7610.2010.02299.x

*McBride-Chang, C., Lam, F., Lam, C., Doo, S., Wong, S. W. L., & Chow, Y. Y. Y. (2008). Word recognition and cognitive profiles of Chinese pre-school children at risk for dyslexia through language delay or familial history of dyslexia. *Journal of Child Psychology and Psychiatry, 49,* 211–218.

McBride-Chang, C., Lin, D., Liu, P. D., Aram, D., Levin, I., Cho, J.-R., . . . Zhang, Y. (2012). The ABC's of Chinese: Maternal mediation of Pinyin for Chinese children's early literacy skills. *Reading and Writing, 25,* 283–300. http://dx.doi.org/10.1007/s11145-010-9270-y

*McBride-Chang, C., Liu, P. D., Wong, T., Wong, A., & Shu, H. (2012). Specific reading difficulties in Chinese, English, or both: Longitudinal markers of phonological awareness, morphological awareness, and RAN in Hong Kong Chinese children. *Journal of Learning Disabilities, 45,* 503–514. http://dx.doi.org/10.1177/0022219411400748

McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing, 8,* 323–339. http://dx.doi.org/10.1007/BF00395112

*McBride-Chang, C., Shu, H., Chan, W., Wong, T., Wong, A. M. Y., Zhang, Y. P., . . . Chan, P. (2013). Poor readers of Chinese and English: Overlap, stability, and longitudinal correlates. *Scientific Studies of Reading, 17,* 57–70. http://dx.doi.org/10.1080/10888438.2012.689787

McBride-Chang, C., Shu, H., Ng, J. Y. W., Meng, X., & Penney, T. (2007). Morphological structure awareness, vocabulary, and reading. In R. K. Wagner, K. Tannenbaum, & A. Muse (Eds.), *Vocabulary development and its implications for reading comprehension* (pp. 104–122). New York, NY: Guilford Press.

*McBride-Chang, C., Shu, H., Zhou, A. B., Wat, C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition. *Journal of Educational Psychology, 95,* 743–751. http://dx.doi.org/10.1037/0022-0663.95.4.743

*McBride-Chang, C., Wagner, R. K., Muse, A., Chow, B. W-Y., & Shu, H. (2005). The role of morphological awareness in children's vocabulary acquisition in English. *Applied Psycholinguistics, 26,* 415–435. http://dx.doi.org/10.1017/S014271640505023X

*McCutchen, D., Green, L., & Abbott, R. D. (2008). Children's morphological knowledge: Links to literacy. *Reading Psychology, 29,* 289–314. http://dx.doi.org/10.1080/02702710801982050

*McCutchen, D., & Logan, B. (2011). Inside incidental word learning: Children's strategic use of morphological information to infer word meanings. *Reading Research Quarterly, 46,* 334–349. http://dx.doi.org/10.1002/RRQ.003

Melby-Lervåg, M., Lyster, S. A., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138,* 322–352. http://dx.doi.org/10.1037/a0026744

*Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40,* 665–681. http://dx.doi.org/10.1037/0012-1649.40.5.665

*Nagy, W., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology, 98,* 134–147. http://dx.doi.org/10.1037/0022-0663.98.1.134

*Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology, 95,* 730–742. http://dx.doi.org/10.1037/0022-0663.95.4.730

Newman, E. H., Tardif, T., Huang, J., & Shu, H. (2011). Phonemes matter: The role of phoneme-level awareness in emergent Chinese readers. *Journal of Experimental Child Psychology, 108,* 242–259. http://dx.doi.org/10.1016/j.jecp.2010.09.001

Nunes, T., & Hatano, G. (2004). Morphology, reading and spelling: Looking across languages. In T. Nunes & P. Bryant (Eds.), *Handbook of children's literacy* (pp. 651–672). Boston, MA: Kluwer Academic Publisher. http://dx.doi.org/10.1007/978-94-017-1731-1_34

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach.* Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511486821

Pan, J., McBride-Chang, C., Shu, H., Liu, H., Zhang, Y., & Li, H. (2011). What is in the naming? A 5-year longitudinal study of early rapid naming and phonological sensitivity in relation to subsequent reading skills in both native Chinese and English as a second language. *Journal of Educational Psychology, 103,* 897–908. http://dx.doi.org/10.1037/a0024344

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, UK: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch13

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112,* 43–59. http://dx.doi.org/10.1037/0033-295X.112.1.43

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115. http://dx.doi.org/10.1037/0033-295X.103.1.56

Reed, D. K. (2008). A synthesis of morphology interventions and effects on reading outcomes for students in Grades K–12. *Learning Disabilities Research & Practice, 23,* 36–49. http://dx.doi.org/10.1111/j.1540-5826.2007.00261.x

*Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in Grades 4, 6, and 8. *Journal of Experimental Child Psychology, 102,* 96–113. http://dx.doi.org/10.1016/j.jecp.2008.01.004

*Saiegh-Haddad, E., & Geva, E. (2008). Morphological awareness, phonological awareness, and reading in English–Arabic bilingual children. *Reading and Writing, 21,* 481–504. http://dx.doi.org/10.1007/s11145-007-9074-x

Savage, R. S., Frederickson, N., Goodwin, R., Patni, U., Smith, N., & Tuersley, L. (2005). Relationships among rapid digit naming, phonological processing, motor automaticity, and speech perception in poor, average, and good readers and spellers. *Journal of Learning Disabilities, 38,* 12–28. http://dx.doi.org/10.1177/00222194050380010201

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & C. A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–119). Timonium, MD: York Press.

Seidenberg, M. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, B. Miller, J. Lee, & O. Tzeng (Eds.), *Dyslexia across languages. Orthography and the brain-gene-behavior link* (pp. 151–174). Baltimore, MD: Brookes.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. http://dx.doi.org/10.1037/0033-295X.96.4.523

Seymour, P. H. (2005). Early reading development in European orthographies. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 296–315). Oxford, UK: Blackwell. http://dx.doi.org/10.1002/9780470757642.ch16

*Shankweiler, D., Crain, S., Katz, L., Fowler, A., Liberman, A., Brady, S., . . . Fletcher, J. (1995). Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax. *Psychological Science, 6,* 149–156. http://dx.doi.org/10.1111/j.1467-9280.1995.tb00324.x

*Shankweiler, D., Lundquist, E., Dreyer, L. G., & Dickinson, C. C. (1996). Reading and spelling difficulties in high school students: Causes and consequences. *Reading and Writing, 8,* 267–294. http://dx.doi.org/10.1007/BF00420279

Shu, H. (2003). Chinese writing system and learning to read. *International Journal of Psychology, 38,* 274–285. http://dx.doi.org/10.1080/00207590344000060

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development, 74,* 27–47. http://dx.doi.org/10.1111/1467-8624.00519

*Shu, H., McBride-Chang, C., Wu, S., & Liu, H. Y. (2006). Understanding Chinese developmental dyslexia: Morphological awareness as a core cognitive construct. *Journal of Educational Psychology, 98,* 122–133. http://dx.doi.org/10.1037/0022-0663.98.1.122

Shu, H., Peng, H., & McBride-Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science, 11,* 171–181. http://dx.doi.org/10.1111/j.1467-7687.2007.00654.x

*Siegel, L. S. (2008). Morphological awareness skills of English language learners and children with dyslexia. *Topics in Language Disorders, 28,* 15–27. http://dx.doi.org/10.1097/01.adt.0000311413.75804.60

*Singson, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading and Writing, 12,* 219–252. http://dx.doi.org/10.1023/A:1008196330239

Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology, 37,* 886–899. http://dx.doi.org/10.1037/0012-1649.37.6.886

Song, S., Georgiou, G. K., Su, M., & Hua, S. (2016). How well do phonological awareness and rapid automatized naming correlate with Chinese reading accuracy and fluency? A meta-analysis. *Scientific Studies of Reading, 20,* 99–123. http://dx.doi.org/10.1080/10888438.2015.1088543

*Song, S., Su, M., Kang, C., Liu, H., Zhang, Y., McBride-Chang, C., . . . Shu, H. (2015). Tracing children's vocabulary development from preschool through the school-age years: An 8-year longitudinal study. *Developmental Science, 18,* 119–131. http://dx.doi.org/10.1111/desc.12190

*Swank, L. K. (1997). Linguistic influences on the emergence of written word decoding in first grade. *American Journal of Speech-Language Pathology, 6,* 62–66. http://dx.doi.org/10.1044/1058-0360.0604.62

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of

the correlation evidence. *Review of Educational Research, 73,* 407–440. http://dx.doi.org/10.3102/00346543073004407

*Tolchinsky, L., Levin, I., Aram, D., & McBride-Chang, C. (2012). Building literacy in alphabetic, abjad and morphosyllabic systems. *Reading and Writing, 25,* 1573–1598. http://dx.doi.org/10.1007/s11145-011-9334-7

*Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology, 103,* 523–534. http://dx.doi.org/10.1037/a0023495

*Tong, X. L., & McBride-Chang, C. (2010). Chinese-English biscriptal reading: Cognitive component skills across orthographies. *Reading and Writing, 23,* 293–310. http://dx.doi.org/10.1007/s11145-009-9211-9

*Tong, X. L., McBride-Chang, C., Shu, H., & Wong, A. M. Y. (2009). Morphological awareness, orthographic knowledge, and spelling errors: Keys to understanding early Chinese literacy acquisition. *Scientific Studies of Reading, 13,* 426–452. http://dx.doi.org/10.1080/10888430903162910

*Tong, X. L., McBride-Chang, C., Wong, A. M. Y., Shu, H., Reitsma, P., & Rispens, J. (2011). Longitudinal predictors of very early Chinese literacy acquisition. *Journal of Research in Reading, 34,* 315–332. http://dx.doi.org/10.1111/j.1467-9817.2009.01426.x

Tong, X., Tong, X., & McBride, C. (2017). Unpacking the relation between phonological awareness and Chinese word reading: Levels of morphological awareness and vocabulary. *Contemporary Educational Psychology, 48,* 167–178. http://dx.doi.org/10.1016/j.cedpsych.2016.07.003

*Tong, X. H., Tong, X. L., Shu, H., Chan, S. F., & McBride-Chang, C. (2014). Discourse-level reading comprehension in Chinese children: What is the role of syntactic awareness? *Journal of Research in Reading, 37,* S48–S70. http://dx.doi.org/10.1111/1467-9817.12016

Treiman, R., & Kessler, B. (2014). *How children learn to write words.* New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199907977.001.0001

Venezky, R. (1967). English orthography: Its graphical structure and its relation to sound. *Reading Research Quarterly, 2,* 75–105. http://dx.doi.org/10.2307/747031

*Wang, M., Cheng, C. X., & Chen, S. W. (2006). Contribution of morphological awareness to Chinese-English biliteracy acquisition. *Journal of Educational Psychology, 98,* 542–553. http://dx.doi.org/10.1037/0022-0663.98.3.542

*Wang, M., Yang, C., & Cheng, C. (2009). The contributions of phonology, orthography, and morphology in Chinese-English biliteracy acquisition. *Applied Psycholinguistics, 30,* 291–314. http://dx.doi.org/10.1017/S0142716409090122

*Wang, Y., McBride-Chang, C., & Chan, S. F. (2014). Correlates of Chinese kindergarteners' word reading and writing: The unique role of copying skills? *Reading and Writing, 27,* 1281–1302. http://dx.doi.org/10.1007/s11145-013-9486-8

*Wang, Y., Yin, L., & McBride, C. (2015). Unique predictors of early reading and writing: A one-year longitudinal study of Chinese kindergarteners. *Early Childhood Research Quarterly, 32,* 51–59. http://dx.doi.org/10.1016/j.ecresq.2015.02.004

*Wei, T. Q., Bi, H. Y., Chen, B. G., Liu, Y., Weng, X. C., & Wydell, T. N. (2014). Developmental changes in the role of different metalinguistic awareness skills in Chinese reading acquisition from preschool to third grade. *PLoS ONE, 9,* e96240. http://dx.doi.org/10.1371/journal.pone.0096240

*Wong, A. M. Y., Ho, C. S. H., Au, T. K. F., Kidd, J. C., Ng, A. K. H., Yip, L. P. W., & Lam, C. C. C. (2015). (Dis)connections between specific language impairment and dyslexia in Chinese. *Reading and Writing, 28,* 699–719. http://dx.doi.org/10.1007/s11145-015-9546-3

*Wong, A. M. Y., Kidd, J. C., Ho, C. S. H., & Au, T. K. F. (2010). Characterizing the overlap between SLI and dyslexia in Chinese: The role of phonology and beyond. *Scientific Studies of Reading, 14,* 30–57. http://dx.doi.org/10.1080/10888430903242043

Wu, N., Zhou, X., & Shu, H. (1999). Sublexical processing in reading Chinese: A development study. *Language and Cognitive Processes, 14,* 503–524. http://dx.doi.org/10.1080/016909699386176

*Xue, J., Shu, H., Li, H., Li, W., & Tian, X. (2013). The stability of literacy-related cognitive contributions to Chinese character naming and reading fluency. *Journal of Psycholinguistic Research, 42,* 433–450. http://dx.doi.org/10.1007/s10936-012-9228-0

Yang, J., Shu, H., McCandliss, B. D., & Zevin, J. D. (2013). Orthographic influences on division of labor in learning to read Chinese and English: Insights from computational modeling. *Bilingualism: Language and Cognition, 16,* 354–366. http://dx.doi.org/10.1017/S1366728912000296

*Yeung, P. S., Ho, C. S. H., Chan, D. W. O., & Chung, K. K. H. (2014). What are the early indicators of persistent word reading difficulties among Chinese readers in elementary grades? *Dyslexia: An International Journal of Research and Practice, 20,* 119–145. http://dx.doi.org/10.1002/dys.1471

*Yeung, P. S., Ho, C. S. H., Chik, P. P. M., Lo, L. Y., Luan, H., Chan, D. W. O., & Chung, K. K. H. (2011). Reading and spelling Chinese among beginning readers: What skills make a difference? *Scientific Studies of Reading, 15,* 285–313. http://dx.doi.org/10.1080/10888438.2010.482149

*Yeung, P. S., Ho, C. S. H., Wong, Y. K., Chan, D. W. O., Chung, K. K. H., & Lo, L. Y. (2013). Longitudinal predictors of Chinese word reading and spelling among elementary grade students. *Applied Psycholinguistics, 34,* 1245–1277. http://dx.doi.org/10.1017/S0142716412000239

Zhang, H. (in press). Development of morphological awareness in in young Chinese readers: Comparing poor comprehenders and good comprehenders. *Reading & Writing Quarterly.* http://dx.doi.org/10.1080/10573569.2016.1145562

*Zhang, J., & McBride-Chang, C. (2014). Auditory sensitivity, speech perception, L1 Chinese, and L2 English reading abilities in Hong Kong Chinese children. *Developmental Psychology, 50,* 1001–1013. http://dx.doi.org/10.1037/a0035086

*Zhang, J., McBride-Chang, C., Tong, X. L., Wong, A. M. Y., Shu, H., & Fong, C. Y. C. (2012). Reading with meaning: The contributions of meaning-related variables at the word and subword levels to early Chinese reading comprehension. *Reading and Writing, 25,* 2183–2203. http://dx.doi.org/10.1007/s11145-011-9353-4

*Zhang, J., McBride-Chang, C., Wong, A. M. Y., Tardif, T., Shu, H., & Zhang, Y. P. (2014). Longitudinal correlates of reading comprehension difficulties in Chinese children. *Reading and Writing, 27,* 481–501. http://dx.doi.org/10.1007/s11145-013-9453-4

Zhang, Y., Tardif, T., Shu, H., Li, H., Liu, H., McBride-Chang, C., . . . Zhang, Z. (2013). Phonological skills and vocabulary knowledge mediate socioeconomic status effects in predicting reading outcomes for Chinese children. *Developmental Psychology, 49,* 665–671. http://dx.doi.org/10.1037/a0028612

*Zhou, Y. L., McBride-Chang, C., Fong, C. Y. C., Wong, T. T. Y., & Cheung, S. K. (2012). A comparison of phonological awareness, lexical compounding, and homophone training for Chinese word reading in Hong Kong kindergartners. *Early Education and Development, 23,* 475–492. http://dx.doi.org/10.1080/10409289.2010.530478

*Zhou, Y., McBride-Chang, C., Law, A. B. Y., Li, T., Cheung, A. C. Y., Wong, A. M. Y., & Shu, H. (2014). Development of reading-related skills in Chinese and English among Hong Kong Chinese children with and without dyslexia. *Journal of Experimental Child Psychology, 122,* 75–91. http://dx.doi.org/10.1016/j.jecp.2013.12.003

Appendix A

Information on the Studies Included in the Meta-Analysis

Study	Language	Reading ability	Grade level	Sample size	Type of PA task	Type of MA task I	Type of MA task II	Type of MA task III	Reading accuracy	Reading fluency	Reading comprehension
Apel & Lawrence,	EN	Blank	Beginning	44	Blank				.61		
2011	EN	Blank	Beginning	44		Prod.	Oral	—	.52		
	EN	Normal	Beginning	44	Blank				.44		
	EN	Normal	Beginning	44		Prod.	Oral	—	.56		
Apel & Thomas-Tate,	EN	Unselct.	Interm.	30	Blank				.38		.02
2009	EN	Unselct.	Interm.	30		Prod.	Blank	—	.65		.21
Apel et al., 2012	EN	Unselct.	Beginning	26	Simple				−.19		−.02
	EN	Unselct.	Beginning	26		Prod.	Blank	—	.60		.71
	EN	Unselct.	Interm.	30	Simple				.03		.12
	EN	Unselct.	Interm.	30		Prod.	Blank	—	.13		.21
Berninger et al., 2006	EN	Poor	Blank	122	Blank				.72	.45	
	EN	Poor	Blank	122		Blank	Blank	—	.76	.53	
Carlisle &	EN	Unselct.	Preschool	101	Blank				.58		
Nomanbhoy, 1993	EN	Unselct.	Preschool	101		Ident.	Oral	—	.27		
	EN	Unselct.	Preschool	101		Prod.	Oral	—	.46		
Chen et al., 2009	CH	Unselct.	Beginning	59	Simple				.37		
	CH	Unselct.	Beginning	59	Simple				.34		
	CH	Unselct.	Beginning	59		Ident.	Oral	Comp.	.33		
	CH	Unselct.	Beginning	59		Prod.	Oral	Comp.	.54		
Cheung et al., 2010	CH	Unselct.	Blank	141	Blank				.28		
	CH	Unselct.	Blank	141		Prod.	Oral	Comp.	.41		
Chung & McBride-	CH	Unselct.	Preschool	85	Blank				.47		
Chang, 2011	CH	Unselct.	Preschool	85		Prod.	Oral	Blank	.40		
Cunningham &	EN	Blank	Interm.	74	Complex				.49		.15
Carroll, 2013	EN	Blank	Interm.	74		Prod.	Oral	—	.25		.30
Deacon & Kirby,	EN	Unselct.	Beginning	103	Complex				.67		.71
2004	EN	Unselct.	Beginning	103		Prod.	Oral	—	.65		.70
Deacon et al., 2007	EN	Unselct.	Beginning	76	Simple				.43		
	EN	Unselct.	Beginning	76		Prod.	Oral	—	.51		
Deacon et al., 2013	EN	Unselct.	Beginning	99	Blank				.70		
	EN	Unselct.	Beginning	99		Prod.	Oral	—	.53		
	EN	Unselct.	Beginning	99		Prod.	Oral	—	.49		
Deacon, 2012	EN	Unselct.	Beginning	123	Blank				.73		
	EN	Unselct.	Beginning	123		Prod.	Oral	—	.51		
	EN	Unselct.	Interm.	79	Blank				.67		
	EN	Unselct.	Interm.	79		Prod.	Oral	—	.40		
Farran et al., 2012	EN	Normal	Blank	83	Blank				.48	.29	
	EN	Normal	Blank	83		Ident.	Oral	—	.17	.03	
Fraser & Conti-	EN	Blank	Blank	71	Complex				.70	.67	.29
Ramsden, 2008	EN	Blank	Blank	71	Simple				.64	.49	.32
	EN	Blank	Blank	71	Simple				.57	.48	.31
	EN	Blank	Blank	71		Prod.	Blank	—	.37	.23	.52
Ho et al., 2011	CH	Poor	Preschool	101	Simple				.51		
	CH	Poor	Preschool	101		Prod.	Oral	Comp.	.35		
Hu, 2013	CH	Unselct.	Interm.	106	Complex				.28		
	CH	Unselct.	Interm.	106	Complex				.40		
	CH	Unselct.	Interm.	106		Prod.	Oral	Comp.	.33		
Jarmulowicz et al.,	EN	Normal	Interm.	76	Blank				.59		.61
2008, 2007	EN	Normal	Interm.	76		Ident.	Oral	—	.52		.55
Joanisse et al., 2000	EN	Poor	Interm.	61	Complex				.30		
	EN	Poor	Interm.	61		Prod.	Oral	—	.27		
	EN	Poor	Interm.	61		Prod.	Oral	—	.30		
	EN	Normal	Interm.	40	Complex				.36		
	EN	Normal	Interm.	40		Prod.	Oral	—	.17		
	EN	Normal	Beginning	37	Complex				.31		
	EN	Normal	Beginning	37		Prod.	Oral	—	.20		

(Appendices continue)

Appendix A (*continued*)

Study	Language	Reading ability	Grade level	Sample size	Type of PA task	Type of MA task I	Type of MA task II	Type of MA task III	Reading accuracy	Reading fluency	Reading comprehension
	EN	Normal	Beginning	37		Prod.	Oral	—	.26		
Kim, Apel, & Al Otaiba, 2013	EN	Poor	Beginning	304	Simple				.45		
	EN	Poor	Beginning	304	Blank				.62		
	EN	Poor	Beginning	304		Prod.	Oral	—	.59		
Kirby et al., 2012	EN	Unselct.	Beginning	103	Blank				.67	.56	.66
	EN	Unselct.	Beginning	103		Prod.	Oral	—	.20	.06	.07
Kruk & Bergman, 2013	EN	Blank	Beginning	157	Blank				.70		.70
	EN	Blank	Beginning	157		Ident.	Oral	—	.53		.54
	EN	Blank	Beginning	157		Prod.	Oral	—	.56		.62
Lam et al., 2008	CH	Poor	Preschool	80	Simple				.39		
	CH	Poor	Preschool	80	Complex				.35		
	CH	Poor	Preschool	80		Prod.	Oral	Comp.	.33		
Lei et al., 2011	CH	Unselct.	Preschool	261	Simple				.35	.32	
	CH	Unselct.	Preschool	261		Ident.	Oral	Comp.	.25	.17	
	CH	Unselct.	Preschool	261		Prod.	Oral	Comp.	.31	.29	
Li & Wu, 2015	CH	Normal	Beginning	135	Simple					.27	.35
	CH	Normal	Beginning	135	Complex					.25	.31
	CH	Normal	Beginning	135		Prod.	Oral	Hmph.		.42	.31
	CH	Normal	Beginning	135		Prod.	Oral	Comp.		.40	.32
	CH	Normal	Beginning	135		Prod.	Oral	Hmgr.		.52	.46
	CH	Normal	Interm.	142	Simple					.26	.27
	CH	Normal	Interm.	142	Complex					.30	.30
	CH	Normal	Interm.	142		Prod.	Oral	Hmph.		.29	.38
	CH	Normal	Interm.	142		Prod.	Oral	Comp.		.33	.28
	CH	Normal	Interm.	142		Prod.	Oral	Hmgr.		.41	.46
	CH	Normal	Interm.	138	Simple					−.08	−.05
	CH	Normal	Interm.	138	Complex					−.07	.13
	CH	Normal	Interm.	138		Prod.	Oral	Hmph.		.07	.18
	CH	Normal	Interm.	138		Prod.	Oral	Comp.		.16	.39
	CH	Normal	Interm.	138		Prod.	Oral	Hmgr.		.21	.40
Li et al., 2012	CH	Unselct.	Preschool	184	Simple				.07		
	CH	Unselct.	Preschool	184	Simple				.45		
	CH	Unselct.	Preschool	184		Ident.	Oral	Hmph.	.12		
	CH	Unselct.	Preschool	184		Prod.	Oral	Comp.	.35		
	CH	Unselct.	Blank	273	Simple				.32		
	CH	Unselct.	Blank	273	Complex				.19		
	CH	Unselct.	Blank	273		Ident.	Oral	Hmph.	.39		
	CH	Unselct.	Blank	273		Prod.	Oral	Hmgr.	.34		
Lin et al., 2012	CH	Unselct.	Preschool	63	Simple				.40		
	CH	Unselct.	Preschool	63		Prod.	Oral	Comp.	.24		
	CH	Unselct.	Preschool	43	Simple				.33		
	CH	Unselct.	Preschool	43		Prod.	Oral	Comp.	.45		
Liu & McBride-Chang, 2014	CH	Unselct.	Interm.	121	Blank				.32		
	CH	Unselct.	Interm.	121		Prod.	Oral	Blank	.56		
Liu et al., 2014	CH	Unselct.	Beginning	50	Blank				.33		
	CH	Unselct.	Beginning	50		Prod.	Oral	Comp.	.44		
	CH	Unselct.	Interm.	50	Blank				.49		
	CH	Unselct.	Interm.	50		Prod.	Oral	Comp.	.53		
Liu et al., 2015	CH	Unselct.	Interm.	92	Blank				.26		.20
	CH	Unselct.	Interm.	92		Prod.	Oral	Comp.	.43		.29
Mahony et al., 2000	EN	Normal	Blank	101	Complex				.57		
	EN	Normal	Blank	101	Complex				.65		
	EN	Normal	Blank	101		Prod.	Written	—	.61		
	EN	Normal	Blank	101		Prod.	Blank	—	.53		

(*Appendices continue*)

Appendix A (*continued*)

Study	Language	Reading ability	Grade level	Sample size	Type of PA task	Type of MA task I	Type of MA task II	Type of MA task III	Reading accuracy	Reading fluency	Reading comprehension
McBride-Chang et al., 2003	CH	Unselct.	Preschool	100	Simple				.36		
	CH	Unselct.	Preschool	100	Complex				.16		
	CH	Unselct.	Preschool	100		Ident.	Oral	Hmph.	.41		
	CH	Unselct.	Preschool	100		Prod.	Oral	Comp.	.52		
	CH	Unselct.	Beginning	100	Simple				.14		
	CH	Unselct.	Beginning	100	Complex				.33		
	CH	Unselct.	Beginning	100		Ident.	Oral	Hmph.	.17		
	CH	Unselct.	Beginning	100		Prod.	Oral	Comp.	.40		
McBride-Chang, Cho, et al., 2005	EN	Unselct.	Beginning	105	Blank				.43		
	EN	Unselct.	Beginning	105		Prod.	Oral	—	.10		
	CH	Unselct.	Beginning	100	Blank				.30		
	CH	Unselct.	Beginning	100		Prod.	Oral	Comp.	.39		
McBride-Chang, Wagner et al., 2005b	EN	Unselct.	Preschool	115	Blank				.58		
	EN	Unselct.	Preschool	115		Ident.	Oral	—	.34		
	EN	Unselct.	Preschool	115		Prod.	Oral	—	.40		
	EN	Unselct.	Beginning	105	Blank				.48		
	EN	Unselct.	Beginning	105		Ident.	Oral	—	.26		
	EN	Unselct.	Beginning	105		Prod.	Oral	—	.18		
McBride-Chang et al., 2006	CH	Unselct.	Preschool	217	Simple				.49		
	CH	Unselct.	Preschool	217	Complex				.19		
	CH	Unselct.	Preschool	217		Ident.	Oral	Comp.	.18		
	CH	Unselct.	Preschool	217		Prod.	Oral	Comp.	.37		
McBride-Chang et al., 2008	CH	Poor	Preschool	72	Simple				.43		
	CH	Poor	Preschool	72	Complex				.38		
	CH	Poor	Preschool	72		Prod.	Oral	Comp.	.46		
McBride-Chang et al., 2011	CH	Poor	Preschool	47	Simple				.37		
	CH	Poor	Preschool	47	Complex				.20		
	CH	Poor	Preschool	47		Prod.	Oral	Comp.	.55		
McCutchen & Logan, 2011	EN	Unselct.	Advanced	88	Blank				.47		.32
	EN	Unselct.	Advanced	88		Prod.	Written	—	.63		.61
	EN	Unselct.	Advanced	88		Ident.	Written	—	.20		.43
	EN	Unselct.	Advanced	88		Ident.	Written	—	.38		.38
	EN	Unselct.	Advanced	74	Blank				.27		.14
	EN	Unselct.	Advanced	74		Prod.	Written	—	.40		.32
	EN	Unselct.	Advanced	74		Ident.	Written	—	.40		.40
	EN	Unselct.	Advanced	74		Ident.	Written	—	.31		.43
McCutchen et al., 2008	EN	Normal	Blank	72	Complex				.43		.23
	EN	Normal	Blank	72		Blank	Blank	—	.09		.16
	EN	Normal	Blank	72		Blank	Blank	—	.37		.41
Muter et al., 2004	EN	Unselct.	Beginning	90	Simple				.42		
	EN	Unselct.	Beginning	90	Simple				.28		
	EN	Unselct.	Beginning	90	Simple				.47		
	EN	Unselct.	Beginning	90	Simple				.42		
	EN	Unselct.	Beginning	90	Complex				.60		
	EN	Unselct.	Beginning	90	Complex				.55		
	EN	Unselct.	Beginning	90		Prod.	Oral	—	.32		
Nagy et al., 2003	EN	Poor	Beginning	98	Blank				.47	.14	.38
	EN	Poor	Beginning	98		Ident.	Blank	—	.22	−.04	.48
	EN	Poor	Beginning	98		Ident.	Blank	—	.18	.05	.46
	EN	Poor	Beginning	98		Ident.	Blank	—	.16	.11	.20
	EN	Poor	Interm.	97	Blank				.59	−.41	.51
	EN	Poor	Interm.	97		Ident.	Blank	—	.43	−.21	.47
	EN	Poor	Interm.	97		Ident.	Blank	—	.32	−.20	.47
	EN	Poor	Interm.	97		Ident.	Blank	—	.22	−.16	.39
Nagy et al., 2006	EN	Unselct.	Blank	182	Complex				.69	.61	.38
	EN	Unselct.	Blank	182		Blank	Blank	—	.67	.48	.76
	EN	Unselct.	Advanced	218	Complex				.65	.64	.53
	EN	Unselct.	Advanced	218		Blank	Blank	—	.65	.61	.65
	EN	Unselct.	Advanced	207	Complex				.53	.59	.53

(*Appendices continue*)

Appendix A (*continued*)

Study	Language	Reading ability	Grade level	Sample size	Type of PA task	Type of MA task I	Type of MA task II	Type of MA task III	Reading accuracy	Reading fluency	Reading comprehension
	EN	Unselct.	Advanced	207		Blank	Blank	—	.50	.55	.59
Roman et al., 2009	EN	Unselct.	Blank	92	Complex				.48		
	EN	Unselct.	Blank	92		Prod.	Oral	—	.64		
Saiegh-Haddad & Geva, 2008	EN	Unselct.	Blank	43	Blank				.55	.53	
	EN	Unselct.	Blank	43		Ident.	Oral	—	.41	.40	
	EN	Unselct.	Blank	43		Ident.	Oral	—	.43	.59	
Shankweiler et al., 1995	EN	Unselct.	Blank	353	Complex				.76		.66
	EN	Unselct.	Blank	353		Prod.	Oral	—	.70		.71
Shankweiler et al., 1996	EN	Blank	Advanced	65	Complex				.53		
	EN	Blank	Advanced	65		Prod.	Oral	—	.46		
Shu et al., 2006	CH	Poor	Advanced	75	Complex				.24		.24
	CH	Poor	Advanced	75	Blank				.34		.39
	CH	Poor	Advanced	75		Prod.	Oral	Hmgr.	.39		.20
	CH	Poor	Advanced	75		Ident.	Oral	Hmph.	.25		.25
	CH	Normal	Advanced	77	Complex				.46		.39
	CH	Normal	Advanced	77	Blank				.47		.54
	CH	Normal	Advanced	77		Prod.	Oral	Hmgr.	.64		.50
	CH	Normal	Advanced	77		Ident.	Oral	Hmph.	.27		.04
Siegel, 2008	EN	Blank	Advanced	1,238	Blank				.43	.47	.29
	EN	Blank	Advanced	1,238		Ident.	Written	—	.47	.49	.46
	EN	Blank	Advanced	1,238		Ident.	Written	—	.46	.46	.43
Swank, 1997	EN	Unselct.	Preschool	60	Complex				.56		
	EN	Unselct.	Preschool	60	Simple				.67		
	EN	Unselct.	Preschool	60	Simple				.49		
	EN	Unselct.	Preschool	60	Simple				.41		
	EN	Unselct.	Preschool	60	Simple				.29		
	EN	Unselct.	Preschool	60	Simple				.37		
	EN	Unselct.	Preschool	60		Blank	Blank	—	.35		
	EN	Unselct.	Preschool	60		Prod.	Blank	—	.48		
Tolchinsky et al., 2012	CH	Unselct.	Preschool	63	Simple				.36		
	CH	Unselct.	Preschool	63	Complex				.04		
	CH	Unselct.	Preschool	63	Complex				.08		
	CH	Unselct.	Preschool	63		Prod.	Oral	Comp.	.25		
Tong et al., 2011	CH	Unselct.	Preschool	187	Simple				.40		
	CH	Unselct.	Preschool	187		Ident.	Oral	Hmph.	.14		
	CH	Unselct.	Preschool	187		Prod.	Oral	Comp.	.45		
Wang et al., 2006	CH	Unselct.	Blank	64	Simple				−.11		.14
	CH	Unselct.	Blank	64	Simple				.35		.20
	CH	Unselct.	Blank	64	Complex				.09		.17
	CH	Unselct.	Blank	64		Ident.	Oral	Comp.	.29		.33
	CH	Unselct.	Blank	64		Prod.	Blank	Comp.	.38		.33
	CH	Unselct.	Blank	64		Ident.	Oral	Hmph.	.11		.21
Wang et al., 2009	CH	Unselct.	Beginning	78	Simple				.30		
	CH	Unselct.	Beginning	78	Simple				.19		
	CH	Unselct.	Beginning	78	Complex				.11		
	CH	Unselct.	Beginning	78		Ident.	Oral	Comp.	.30		
Wang et al., 2014	CH	Unselct.	Preschool	94	Simple				.03		
	CH	Unselct.	Preschool	94		Prod.	Oral	Comp.	.40		
Wang et al., 2015	CH	Unselct.	Preschool	73	Simple				.22		
	CH	Unselct.	Preschool	73		Prod.	Oral	Comp.	.23		
Wei et al., 2014	CH	Unselct.	Preschool	101	Simple				.29		
	CH	Unselct.	Preschool	101	Complex				.45		
	CH	Unselct.	Preschool	101		Prod.	Oral	Comp.	.29		
	CH	Unselct.	Preschool	101		Ident.	Oral	Hmgr.	.29		
	CH	Unselct.	Beginning	94	Simple				.12		
	CH	Unselct.	Beginning	94	Complex				.19		
	CH	Unselct.	Beginning	94		Prod.	Oral	Comp.	.27		
	CH	Unselct.	Beginning	94		Ident.	Oral	Hmgr.	.31		
	CH	Unselct.	Beginning	98	Simple				.42		

(*Appendices continue*)

Appendix A (*continued*)

Study	Language	Reading ability	Grade level	Sample size	Type of PA task	Type of MA task I	Type of MA task II	Type of MA task III	Reading accuracy	Reading fluency	Reading comprehension
	CH	Unselct.	Beginning	98	Complex				.31		
	CH	Unselct.	Beginning	98		Prod.	Oral	Comp.	.32		
	CH	Unselct.	Beginning	98		Ident.	Oral	Hmgr.	.33		
	CH	Unselct.	Interm.	98	Simple				.15		
	CH	Unselct.	Interm.	98	Complex				.30		
	CH	Unselct.	Interm.	98		Prod.	Oral	Comp.	.37		
	CH	Unselct.	Interm.	98		Ident.	Oral	Hmgr.	.19		
Wong et al., 2010	CH	Unselct.	Blank	34	Complex				.33	.41	
	CH	Unselct.	Blank	34	Complex				.13	.25	
	CH	Unselct.	Blank	35		Prod.	Oral	Comp.	.71	.69	
Wong et al., 2015	CH	Unselct.	Preschool	93	Simple				.33	.38	
	CH	Unselct.	Preschool	93	Simple				.24	.28	
	CH	Unselct.	Preschool	92		Ident.	Oral	Hmgr.	.56	.44	
	CH	Unselct.	Preschool	92		Prod.	Oral	Comp.	.59	.48	
Xue et al., 2013	CH	Unselct.	Beginning	408	Blank				.29		
	CH	Unselct.	Beginning	408		Prod.	Oral	Hmgr.	.24		
	CH	Unselct.	Interm.	428	Blank				.31		
	CH	Unselct.	Interm.	428		Prod.	Oral	Hmgr.	.32		
	CH	Unselct.	Advanced	496	Blank				.38		
	CH	Unselct.	Advanced	496		Prod.	Oral	Hmgr.	.27		
Yeung et al., 2011	CH	Unselct.	Beginning	290	Simple				.21		.15
	CH	Unselct.	Beginning	290		Ident.	Oral	Hmph.	.48		.39
Zhang & McBride-Chang, 2014	CH	Unselct.	Blank	153	Blank				.35		
	CH	Unselct.	Blank	153		Prod.	Oral	Comp.	.51		
Zhou et al., 2012	CH	Unselct.	Preschool	88	Simple				.50		
	CH	Unselct.	Preschool	88		Prod.	Oral	Comp.	.38		
	CH	Unselct.	Preschool	88		Prod.	Oral	Hmph.	.13		

Note. PA = phonological awareness; MA = morphological awareness; EN = English; CH = Chinese; Unselct. = unselected; Interm. = intermediate; Prod. = production; Judg. = judgement; Comp. = compounding; Hmph. = homophone; Hmgr. = homograph.

(*Appendices continue*)

Appendix B

Language Comparisons Within Phonological Awareness and Morphological Awareness (top half) and Meta-Linguistic Awareness Comparisons Within English and Chinese (bottom half) of Partial Correlations

	r-values (95% CI)				
	CH	EN	r-values defferences	$Q_{between}$	p
Phonological awareness					
Reading accuracy	.198*** [.148, .247]	.424*** [.375, .471]	EN > CH	40.233	<.001
Reading fluency	.159*** [.079, .236]	.404*** [.339, .465]	EN > CH	22.743	<.001
Reading Comprehension	.145*** [.082, .207]	.235*** [.133, .332]	n.s.	2.177	.140
Morphological awareness					
Reading accuracy	.329*** [.280, .376]	.288*** [.238, .336]	n.s.	1.385	.239
Reading fluency	.239** [.106, .365]	.191** [.077, .300]	n.s.	.302	.583
Reading comprehension	.319*** [.262, .374]	.393*** [.312, .468]	n.s.	2.184	.139
	PA	MA			
Chinese					
Reading accuracy	.198*** [.148, .247]	.329*** [.280, .376]	PA < MA	13.771	<.001
Reading fluency	.159*** [.079, .236]	.239** [.106, .365]	n.s.	1.061	.303
Reading comprehension	.145*** [.082, .207]	.319*** [.262, .374]	PA < MA	16.352	<.001
English					
Reading accuracy	.424*** [.375, .471]	.288*** [.238, .336]	PA > MA	15.154	<.001
Reading fluency	.404*** [.339, .465]	.191** [.077, .300]	PA > MA	11.044	.001
Reading comprehension	.235*** [.133, .332]	.393*** [.312, .468]	PA < MA	6.017	.014

Note. CH = Chinese; EN = English; PA = phonological awareness; MA = morphological awareness.
** $p < .01$. *** $p < .001$.

Received August 28, 2016
Revision received May 15, 2017
Accepted May 18, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 203–217

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000211

Testing Prepares Students to Learn Better: The Forward Effect of Testing in Category Learning

Hee Seung Lee and Dahwi Ahn
Yonsei University

The forward effect of testing occurs when testing on previously studied information facilitates subsequent learning. The present research investigated whether interim testing on initially studied materials enhances the learning of new materials in category learning and examined the metacognitive judgments of such learning. Across the 4 experiments, participants learned the painting styles of various artists, which were divided into 2 separate sections (Sections A and B). They were given an interim test or not on the studied paintings of Section A before moving on to study the paintings of different artists in Section B, and then were given a final test on Section B where participants had to transfer what they had previously learned to new exemplars of the studied artists in Section B. In all experiments, transfer performance on Section B was greater when the participants were given an interim test versus no test. The beneficial effect of interim testing was obtained when the final test was presented in cued-recall (Experiments 1 and 2) and multiple-choice (Experiments 3 and 4) formats. Experiments 3 and 4 also indicated that the forward effect of testing was not due to re-exposure to previously studied items but the testing itself. However, the metacognitive measures provided by the participants did not reflect their actual performance, suggesting that the participants were unaware about the beneficial effects of interim testing. Interim testing appears to prepare students to learn better, facilitating not only learning of specific instances but also generalization of that learning.

Educational Impact and Implications Statement

The present study suggests that an interim test on earlier studied categories can facilitate subsequent learning of new categories. Interim testing appears to prepare students to learn better, facilitating not only learning of specific instances but also generalization of that learning. The results highlight the idea that tests are powerful tools for learning and educators may want to use tests as a preparation for subsequent learning.

Keywords: forward effect of testing, interim testing, category learning, metacognition

A large body of research has shown that testing can enhance student learning (for reviews, see Rawson & Dunlosky, 2012; Roediger & Butler, 2011; Roediger & Karpicke, 2006b). Retrieval of an item from memory can influence the subsequent representation of that item in memory, thus can work as a memory modifier (Bjork, 1975). After taking a test (or retrieval practice) of previously studied information, learners generally perform better on a delayed memory task than those who studied the same material twice without a test (e.g., Butler, 2010; McDaniel & Fisher, 1991; Mcdaniel, Roediger, & McDermott, 2007; Roediger & Karpicke, 2006a). The beneficial effects of testing even occur when no feedback is provided (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Carpenter, 2011; Halamish & Bjork, 2011; Rowland, Littrell-Baez, Sensenig, & DeLosh, 2014) and when retrieval attempts are unsuccessful during a test (e.g., Kornell, Hays, & Bjork, 2009; Hays, Kornell, & Bjork, 2013; Richland, Kornell, & Kao, 2009). This phenomenon, termed as the *testing effect*, suggests an important educational implication in that testing is not just an evaluation instrument of learning but also a powerful tool for learning.

Researchers have also investigated whether testing on previously studied information can enhance subsequent learning, a phenomenon referred to as *test-potentiated learning* (Arnold & McDermott, 2013; Izawa, 1966; Soderstrom & Bjork, 2014), which is distinguished from what we generally call the testing effect. Under typical testing-effect conditions, retrieval practice of previously studied information increases the possibility that the "tested" information will be recalled later. Thus, it can be viewed as what Pastötter and Bäuml (2014) referred to as the *backward effect of testing* or what Roediger and Karpicke (2006b) termed the *direct effect of testing*. In an educational context, however, students can benefit from not just the backward effect of testing, but also a forward effect of testing. For example, the testing experi-

This article was published Online First May 25, 2017.

Hee Seung Lee and Dahwi Ahn, Department of Education, Yonsei University.

This research was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A5A8018731).

Correspondence concerning this article should be addressed to Hee Seung Lee, Department of Education, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Korea. E-mail: hslee00@yonsei.ac.kr

ence may offer students opportunities to evaluate their current study strategies and consider how they can adjust such strategies in the future. In other words, testing may help students become prepared to learn better when provided with an additional opportunity to restudy the same materials or to study newly presented learning materials. In this case, the beneficial effect of testing can be viewed as either the *forward effect of testing* (Pastötter & Bäuml, 2014) or the *indirect effect of testing* (Roediger & Karpicke, 2006b) as the tests promote subsequent learning. Although the forward effect of testing can occur, regardless of whether subsequent learning involves old (i.e., previously presented) items or new materials, the phenomenon termed the *interim-test effect* specifically focuses on the effect of testing on the subsequent learning of "new" information (Szpunar, McDermott, & Roediger, 2008; Wissman, Rawson, & Pyc, 2011). The current study particularly focused on the interim-test effect.

Several studies have illustrated that the testing of previously studied information can enhance the learning of subsequently presented new information (e.g., Weinstein, McDermott, & Szpunar, 2011; Wissman et al., 2011). The interim-test effect has been recently replicated in research with various learning materials, including word lists (Nunes & Weinstein, 2012; Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Szpunar et al., 2008; Wahlheim, 2015; Weinstein, Gilmore, Szpunar, & McDermott, 2014), pictures (Pastötter, Weber, & Bäuml, 2013), videos (Szpunar, Khan, & Schacter, 2013; Yue, Soderstrom, & Bjork, 2015), and faces and names (Weinstein et al., 2011). For example, in the study by Wissman et al. (2011), participants read three sections of complex expository texts and the researchers examined how interim testing of the first two sections influenced the memory of the final third section. Under the interim-test condition, the participants were prompted to recall each preceding section before moving on to the next section, whereas under the no-test condition, the participants were asked to recall the final section only. Across the five experiments in their study, the recall performance of the final section was always greater when the participants were given an interim recall tests. Wissman et al. (2011) suggested that interim testing promotes more effective subsequent encoding strategies. Other researchers have also suggested that interim testing can reduce mind-wandering (Szpunar et al., 2013) and enhance list differentiation by protecting against proactive interference (Szpunar et al., 2008; Wahlheim, 2015).

Although the forward effect of testing appears to have useful implications for education, to the best of our knowledge, all of the existing research on the interim-test effect involves retention tests that focus on how well learners can recall texts, words, names and other related aspects. In other words, such memory tests require learners to simply remember specific instances. However, in an educational context, what is important to learn often goes beyond memorizing specific items. For example, teachers may present two paintings by Claude Monet, such as "Water Lilies" and "In the Garden," as representations of impressionism. In this case, from these examples, the students must not only learn specific instances/episodes but also general characteristics of impressionism that should be abstracted from the paintings. However, if the students fail to abstract the principles or patterns underlying the examples (even if they remember and retain specific instances), then they will be unable to generalize their learning in other circumstances outside the classroom. Only when they can induce and identify abstract patterns from the studied examples will they be able to transfer such knowledge into other examples. This line of reasoning renders *category learning*, which requires students to generalize what they have learned from specific instances into other instances of a certain category. Although category learning is an important aspect of education, previous investigations of such learning have been focusing on testing alternative theories for theory development (for a review, see Murphy, 2004), rather than determining how to optimize the learning of categories (Jacoby, Wahlheim, & Coane, 2010). Despite the extensive theoretical research, limited studies have examined the optimized conditions of category learning (e.g., Carvalho & Goldstone, 2015; Kornell & Bjork, 2008; Jacoby et al., 2010).

Regarding the testing effect in category learning, most previous studies examined a testing condition as a part of categorization training method (e.g., Ashby, Maddox, & Bohil, 2002; Carvalho & Goldstone, 2015). For example, learners studied a series of exemplars either with or without category assignment during a learning phase. The participants who were not given category assignment had to put themselves into a self-testing condition from the beginning of the learning phase rather than after the learning phase. The study that directly examined the testing effect was done by Jacoby et al. (2010). In their study, participants were tested or not regarding their initial learning of natural concepts (bird families) and then were given a final test on the studied categories. The results showed that testing enhanced both recognition memory and classification accuracy compared with the restudy condition. Although they used both studied and novel exemplars of bird families in their final tests, the final test items only included the tested categories. Therefore, the beneficial effect of testing found in Jacoby et al.'s (2010) study can be viewed as a *backward effect* of testing in category learning.

The current study aimed to determine whether the forward effect of testing, with specific focus on the interim-test effect, applies to category learning. To our knowledge, this is the first study that examines whether interim testing on previously studied categories can facilitate subsequent learning of new categories. We had participants learn two groups of learning materials across two separate sections (Sections A and B) and examined whether an interim test on Section A facilitates subsequent learning of Section B. More specifically, the learners initially studied Section A and either did or did not take an interim test on Section A before moving on to Section B. After studying Section B, all the participants were subsequently administered a final transfer test on Section B. The assumption here is that the interim-test effect occurs when the participants tested on Section A perform better on the test regarding Section B. It is important to note that none of the participants were administered a practice test on Section B; that is, it was the first time that the learners were tested on Section B in the final test.

For all the experiments, a painting-style learning task was chosen to examine the interim-test effect in category learning. This task was successfully applied in several previous studies that examined the spacing or interleaving effect in inductive learning (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008; Kornell, Castel, Eich, & Bjork, 2010). In this task, the learners first study a series of paintings by multiple artists and then they are tested with unfamiliar paintings created by the artists who were studied. To successfully perform the task, the learners must first abstract

general patterns of the paintings by the various artists and then transfer what they have learned to novel paintings created by the artists who were studied. Thus, only when the learners study the paintings at a category level (i.e., artist level) will they be able to identify new exemplars of the learned category (i.e., new paintings created by the artists who were studied).

Figure 1 illustrates a schematic representation of the procedures used in the four experiments. In each experiment, participants were either tested or not tested on Section A, after which they moved on to study Section B. In the final test, the participants were either tested on Section B (Experiment 1) or on both Sections A and B (Experiments 2 to 4). Before the final test, the participants also made metacognitive judgments by predicting their own performance on the upcoming final test. Even though the tests are generally known to improve metacognitive judgments (e.g., Baars, Van Gog, De Bruin, & Paas, 2014; King, Zechmeister, & Shaughnessy, 1980; Little & McDaniel, 2015), students are generally unaware of the beneficial effects of testing (for a review, see Karpicke et al., 2009; Kornell & Bjork, 2008). In addition, students often believe that restudy is more effective than testing (e.g., Kornell & Son, 2009). In category learning, however, Jacoby et al. (2010) reported a high correlation between metacognitive judgments and the actual performance in learning of bird families; thus, suggesting that the participants' predictions of their abilities may differ, depending on the type of task and learning materials. Accordingly, the present study measured metacognitive judgments to investigate this issue, along with the interim-test effect in category learning.

Experiment 1 focused on determining whether the interim-test effect occurs in this domain of category learning by comparing the test and no-test groups, while Experiment 2 examined whether it is possible to replicate Experiment 1 when the final transfer test includes categories from both Sections A and B. Experiments 3 and 4 examined whether the beneficial effect of interim testing is because of high levels of initial learning or the testing itself, with regard to both transfer (Experiments 3 and 4) and recognition (only Experiment 4) performance. In describing each of the following experiments, because the main goal of the present study was to examine whether the forward effect of testing applies to category learning, we first focused on reporting learning performance results in each experiment and then discussed metacognitive results of all four experiments in a separate section.

Experiment 1

Method

Participants. In total, 37 undergraduate students (26 women, 11 men; Mean age = 22 years) from a large university in Korea participated in exchange for course credit or monetary compensa-

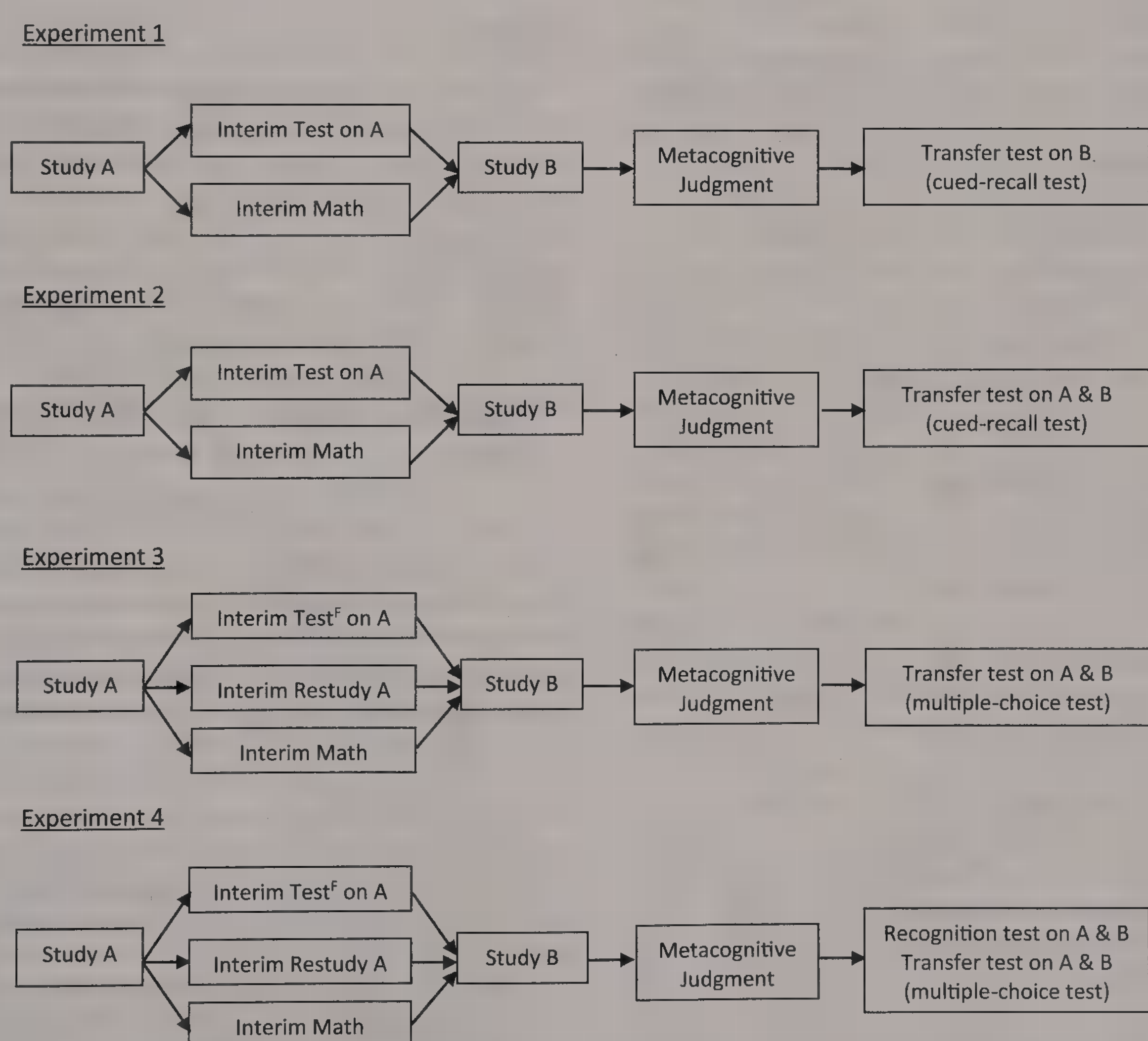

Figure 1. Schematic representation of the procedures used in Experiments 1–4. Test^F represents test with feedback.

tion equivalent to \$5. Sample size for all experiments was determined based on previous studies of interim-test effect (Szpunar et al., 2008, 2013). Each participant was randomly assigned to one of the two conditions, for a total of 19 in the interim-test condition and 18 in the interim-math condition.

Design. A one-way between-subjects design was used to examine the interim-test effect in category learning. Participants were either tested (interim-test condition) or not tested (interim-math condition) on Section A before moving on to studying Section B.

Materials and procedure. The study was conducted according to Human Ethics Guidelines approved by the university where this research took place in Korea. All the participants were individually tested on a computer. In the beginning of the experiment, the participants were informed about the purpose and general procedure of the study. The participants were also told that they would study a series of paintings by 12 different artists across two separate sections (six artists per section) and that their task was to identify the painting style of each artist. More important, all the participants were informed that they would be tested later by being presented with previously unseen paintings that were created by the artists who were studied. The paintings used in the present study were adapted from Kornell and Bjork's (2008) study.[1] All of the paintings were natural landscapes (in color). The paintings were created by 12 different artists (Gorges Braque, Henri-Edmond Cross, Judy Hawkins, Philip Juras, Ryan Lewis, Marilyn Mylrea, Bruno Pessani, Ron Schlorff, Georges Seurat, Emma Ciadi, George Wexler, and Yie Mei). Among the 12 artists, the paintings of six artists were assigned to Section A, while the other six paintings were assigned to Section B. The artist—section pairs were counterbalanced to control for specific item effect.

In Section A, all the participants first studied a total of 36 paintings that consisted of six paintings by each of the six artists. The paintings were presented one at a time in the middle of the computer screen for 3 s, with the last name of the artist displayed below the image (in Korean letters). Because several previous studies have illustrated that inductive learning is enhanced when materials are intermingled, rather than being grouped by the same category (e.g., Kang & Pashler, 2012; Kornell & Bjork, 2008), we presented the paintings of all six artists in a fixed random order. The presentation of each painting was followed by a 0.5-s blank screen. After completing Section A, the participants in the interim-test condition were given a cued-recall test on the section in which they were presented with the same set of paintings (i.e., 36 paintings) from Section A but without the artist's name. They were prompted to enter the artist's name at their own pace and without feedback. In contrast, in the control condition (i.e., the interim-math), the participants were not administered a test on Section A. Instead, they were given a series of simple arithmetic problems. This self-paced, intervening activity was included to address the possibility that the interim-test effect is not because of interim testing per se, but rather the intervening activity (Wissman et al., 2011). In total, 36 problems were presented to equate the number of problems presented in the interim-test and interim-math conditions. Upon completion of either the interim-test or interim-math, the participants continued on to Section B in which they studied another set of paintings created by six different artists. As in Section A, there were a total of 36 paintings that consisted of six paintings by each of the six artists. They were presented in the same manner as in Section A.

Subsequently, the participants made metacognitive judgments about their performance. More specifically, using a number between 0 and 100, they were asked to predict the likelihood of correctly indicating who created the presented unfamiliar paintings, despite the fact that they were created by the same artists who they studied in Section B. The purpose of such judgments was to measure the participants' predictions of their ability to identify novel exemplars from the studied categories. In this regard, such ability can be viewed as *category learning judgment* (Jacoby et al., 2010).[2] Measuring the participants' prediction allowed us to examine the interim-test effect on metacognition in category learning. After providing metacognitive judgments, all the participants were administered a final test, which was a transfer test on Section B. The final test presented a new set of paintings created by the artists who were studied during Section B. There were four paintings by each of the six artists, resulting in a total of 24 paintings. Participants were presented with a painting one at a time in a fixed random order and prompted to enter the artist's name who they believed created the painting. The presentation of each painting was followed by a 0.5-s blank screen. It was a self-paced task and there was no feedback. After completing the test, the participants were debriefed and thanked.

Results

Interim activity performance between Sections A and B. To score participants' answers, we counted only correctly spelled answers of artists' names as correct across all of the reported experiments. Only the participants in the interim-test condition were tested on Section A, and the mean percentage of correct responses was 31.87 (SD = 23.03). Cronbach's α of the interim test was .914. In the interim-math condition (control) the participants solved simple arithmetic problems, and the mean percentage of correct responses was 94.14 (SD = 5.38).

Final test performance. Figure 2 (left) shows the mean percentage of correct responses on the transfer test of Section B in the interim-test and interim-math conditions. Cronbach's α of the final test was .910. An independent t test, conducted on the transfer score of Section B, revealed that the mean difference between the two conditions was statistically significant, $t(35) = 4.35, p < .001$, d = 1.47. The participants who were given an interim-test on Section A (M = 57.68, SD = 20.42) showed significantly better performance on Section B than those in the interim-math condition (M = 25.69, SD = 24.22).

[1] The paintings used in this study are courtesy of Nate Kornell at Williams College, http://sites.williams.edu/nk2/stimuli/. Because the paintings of one artist (Ciprian Stratulat) had low resolution, they were replaced by the paintings of Emma Ciadi.

[2] In Jacoby et al. (2010), category learning judgments (CLJs) were made on each of the studied categories for assessing the testing effects on metacognition at the level of categories. In the present study, metacognitive judgments were made on the groups of studied categories, which were separated by sections. Thus, we continued to use the term *metacognitive judgments* instead of CLJs.

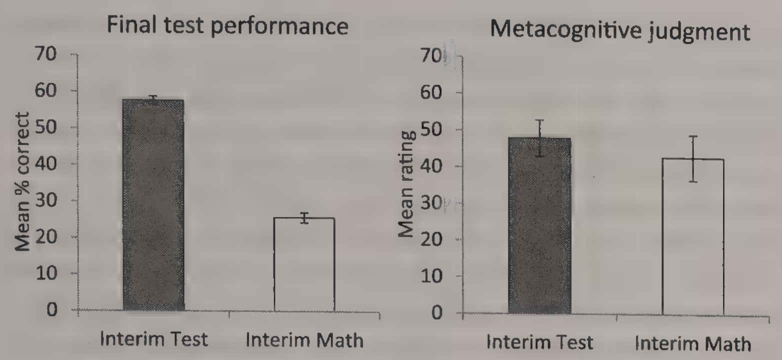

Figure 2. Mean percentage of the correct responses in the final transfer test on Section B (left) and mean ratings of the metacognitive judgments (right) in the interim-test and interim-math conditions of Experiment 1. Error bars represent 1 *SEM*.

Discussion

The results illustrated that interim testing on the previously studied categories facilitated the learning of subsequently presented new categories, as shown in the better transfer performance in the interim-test condition. Note that the participants in the interim-test and interim-math conditions had an equal amount of study time for Section B, and the only difference between the two groups was whether they were given or not given an interim-test on Section A before moving on to Section B. Moreover, all the participants were forewarned that they would be subsequently tested at the commencement of the experiment. Thus, the observed difference between the two conditions was not likely because of the test expectancy effect. In contrast to previous studies on the interim-test effect (e.g., Cho, Neely, Crocco, & Vitrano, 2017; Szpunar et al., 2008; Wissman et al., 2011), the present study did not use a retention task that required the participants to recall specific instances (e.g., words, texts). Instead, a transfer task was applied, which required the participants to apply what they had previously learned to new instances of the studied categories. Thus, the results expand the previous findings on the interim-test effect by demonstrating that interim testing can also facilitate category learning.

Experiment 2

In many educational contexts, a final test administered at the end of a class often deals with the entire materials of the course, rather than a small portion of materials that were not tested in any of the previous quizzes. Therefore, Experiment 2 examined the interim-test effect in a more educationally relevant situation by including all the categories covered during the study for the final transfer test. Similar to Experiment 1, the participants were tested or not tested on Section A before moving on to study Section B, after which the final test included the categories from both Sections A and B. This procedure allowed us to investigate both what we generally call the *testing effect* (backward effect of testing), and the *interim-test effect* (forward effect of testing) in the domain of category learning. If the interim-test group exhibits better transfer performance on Section A, then this can be interpreted as the typical testing effect in category learning. If the interim-test group demonstrates better transfer performance on Section B, then this can be interpreted as the interim-test effect in category learning.

Method

Participants. In total, 30 undergraduate students (14 women, 16 men; Mean age = 22 years) participated in exchange for course credit or $5. Each participant was randomly assigned to one of the two conditions, for a total of 14 in the interim-test condition and 16 in the interim-math condition.

Design, materials, and procedure. As illustrated in Figure 1, the design and procedures used in Experiment 2 were identical to those of Experiment 1, except for two modifications. First, the participants had to provide metacognitive judgments for both Sections A and B. They made their judgments by predicting the likelihood of correctly identifying novel paintings created by the artists who were studied in Section A and Section B. Second, the participants had to complete the final transfer test on the categories from both Sections A and B. Thus, the materials of Experiment 2 required new paintings that the participants had not previously seen during their study, but were created by the artists who were studied from both Sections A and B. The final test included four paintings by each of the 12 artists (six artists per section), resulting in a total of 48 paintings. During the test, the paintings were presented in a fixed random order and they were not separated by sections. In other words, the participants were unaware about which section of artists created each painting. All the other procedures were identical to those presented in Experiment 1.

Results

Interim activity performance between Sections A and B. Only the participants in the interim-test condition were tested on Section A, and the mean percentage of correct responses was 31.94 (SD = 18.77). Cronbach's α of the interim test was .867. In the interim-math condition, the participants solved simple arithmetic problems, and the mean percentage of correct responses was 96.53 (SD = 3.29).

Final test performance. Figure 3 (left) presents the mean percentage of correct responses on the transfer test of Sections A and B in the interim-test and interim-math conditions. Cronbach's α of the final test was .893 on Section A and .886 on Section B, respectively. A 2 $\times$ 2 mixed analysis of variance (ANOVA) was conducted on the number of correct responses. Interim activity (test vs. math) was included as a between-subjects factor, while

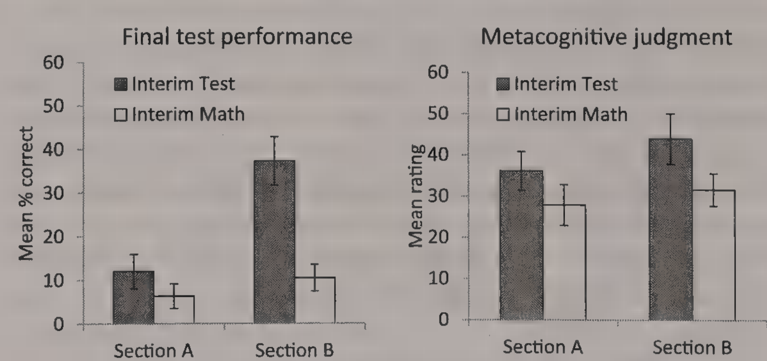

Figure 3. Mean percentage of the correct responses in the final transfer test on Sections A and B (left) and mean ratings of the metacognitive judgments (right) in the interim-test and interim-math conditions of Experiment 2. Error bars represent 1 *SEM*.

section (A vs. B) was included as a within-subject factor. There was a significant main effect of interim activity, $F(1, 28) = 12.35$, $p = .002$, $\eta_p^2 = .306$, such that the interim-test group ($M = 24.56$, $SD = 21.94$) showed significantly better transfer performance than the interim-math group ($M = 8.33$, $SD = 11.83$), regardless of section. The overall effect of the section was also significant, $F(1, 28) = 23.82$, $p < .001$, $\eta_p^2 = .460$, in that the participants correctly identified significantly more paintings when they were from the artists of Section B ($M = 22.92$, $SD = 21.41$) than Section A ($M = 8.89$, $SD = 13.16$), regardless of the interim activity conducted between the two sections. More interestingly, there was a significant interaction effect between interim activity and section, $F(1, 28) = 12.25$, $p = .002$, $\eta_p^2 = .304$, implying that the effect of interim activity differed depending on the section. For Section A, the participants from both conditions showed poor performance (interim-test: $M = 11.90$, $SD = 14.79$; interim-math: $M = 6.25$, $SD = 11.38$); further, the mean performance between the two conditions did not differ, $t(28) = 1.18$, $p = .247$. In contrast, for the Section B, participants who were administered an interim-test on Section A ($M = 37.20$, $SD = 20.90$) showed significantly better transfer performance than those in the interim-math condition ($M = 10.42$, $SD = 12.27$), $t(28) = 4.35$, $p < .001$, $d = 1.64$.

Discussion

Similar to Experiment 1, the results indicated that there was a beneficial effect of interim test in category learning. The participants who were tested on the previously studied categories showed better learning of new categories, as shown in the better transfer performance in the final test regarding these new categories. However, Experiment 2 failed to obtain the typical *testing effect* regarding the categories presented in Section A. The participants in the interim-test and interim-math conditions demonstrated poor transfer performance when they were tested on the categories of Section A (overall $M = 8.89\%$). One possible explanation for this is that there might have been a so-called *floor effect.* In this experiment the participants had to study multiple paintings by 12 different artists, which required them to not only learn the painting styles of the artists but also their names. Considering that the participants were prompted to enter the names of the artists on their own (simply based on their memory in the final test), even if they had successfully learned the painting styles, they might have failed to recall the exact names of the artists. This recall memory was probably worse for Section A, which had a longer time delay before the final test, than for Section B. Indeed, some participants' responses indicated that they had difficulty recalling exact names. A couple of participants wrote the answers like "starts with s" (in Korean) in a consistent manner, indicating that they simply failed to recall the artists' names while knowing the style of paintings. Thus, the following experiments addressed this possibility by assessing the participants' classifications through a multiple-choice test to ensure that the participants are not required to recall the names of the artists in the final transfer test.

Experiment 3

Experiments 1 and 2 demonstrated that interim testing can facilitate learning of a new category. In both experiments, the participants who were tested on the previously studied categories showed better transfer performance regarding the subsequently studied categories. Although it is apparent that interim testing facilitates the subsequent learning of new categories, it is unclear which cognitive process actually promotes such learning. One may argue that the beneficial effect is not because of testing per se but rather the re-exposure to the studied materials (Cho et al., 2017; Kang, McDermott, & Roediger, 2007; Putnam & Roediger, 2013). In Experiments 1 and 2, only the interim-test group was re-exposed to the studied materials, while the interim-math group performed irrelevant interim tasks (i.e., solving math problems). Even though feedback was not provided when the group was tested on Section A, owing to the fact that they were exposed to the materials of Section A twice (i.e., one during the study session and another during the test session), they might have had a better encoding of Section A. As a result, such encoding might have influenced the encoding of subsequently presented materials in Section B. To address this possibility, Experiment 3 included an interim-restudy condition as a comparison group, in addition to the interim-math condition. The interim-restudy group restudied Section A without receiving a test on the section before moving on to Section B. This comparison group would serve as a baseline for evaluating whether the observed interim-test effect is because of the re-exposure or the testing itself. The interim-restudy group is also a more ecologically relevant control group based on the fact that teachers generally do not assign students completely irrelevant interim tasks in the middle of the class simply because they are moving on to a new section.

Furthermore, we modified the test format of the final test from a cued-recall test to a multiple-choice test. In an educational context, although category learning often involves both the learning of category names and the abstraction of category features, we thought a cued-recall test format might have made the task too difficult for some participants. In addition, because the current study mainly aims to investigate whether the interim tests can facilitate category learning (not memory of a specific instance), it will be more appropriate to evaluate whether learners can discriminate between different categories, rather than determining whether they can recall specific names of categories.

Method

Participants. In total, 60 undergraduate students (34 women, 26 men; Mean age = 21 years) participated in exchange for course credit or $5. Each participant was randomly assigned to one of the three conditions, for a total of 20 in the interim-test condition, 19 in the interim-restudy condition, and 21 in the interim-math condition.

Design, materials, and procedure. As illustrated in Figure 1, the overall procedure of Experiment 3 was similar to that of Experiment 2, except for two changes. First, Experiment 3 manipulated the type of interim activity by using three levels: test, restudy, and math. All the participants first studied Section A and subsequently they were administered a different interim activity prior to Section B. In the interim-test condition, the participants were given a cued-recall test on the studied paintings from Section A. After entering the artist's name who they believed had created the presented painting, the participants received an immediate feedback that was presented for 1.5 s. The feedback page simultaneously presented the name of the correct artist with the painting.

In contrast to Experiment 2, this feedback was included to equate the amount of re-exposure to the materials of Section A between the interim-test and interim-restudy conditions. In the interim-restudy condition, the participants were informed that they would be presented again with the studied paintings, and the same materials of the Section A were repeated in a new random order. In the interim-math condition, as in the Experiments 1 and 2, the participants solved simple arithmetic problems.

Second, the final test was presented in a multiple-choice format, and the paintings were identical to those used in the final test of Experiment 2. In each test trial, a painting was presented and the participants made selections among the 12 different artists presented on the page. The names of the artists were alphabetically ordered, which remained unchanged during the entire test session. The test was self-paced, and there was no feedback. All the other procedures were identical to those of Experiment 2.

Results

Interim activity performance between Sections A and B. Only the participants in the interim-test condition were tested on Section A, and the mean percentage of correct responses was 61.67 (SD = 23.34). Cronbach's α of the interim test was .908. In the interim-math condition, the participants solved simple arithmetic problems, and the mean percentage of correct responses was 96.56 (SD = 3.82).

Final test performance. Figure 4 (left) shows the mean percentage of correct responses on the transfer test of Sections A and B in the interim-test, interim-restudy, and interim-math conditions. Cronbach's α of the final test was .830 on Section A, and .865 on Section B, respectively. A 3 × 2 mixed ANOVA was performed on the number of correct responses. Interim activity (test vs. restudy vs. math) was included as a between-subjects factor, while section (A vs. B) was included as a within-subject factor. There was a significant main effect of interim activity, $F(2, 57) = 9.21$, $p < .001$, $\eta_p^2 = .244$. This was because the interim-math group showed worse transfer performance than the other groups, regardless of section. The interim-test group (M = 57.60, SD = 13.06) showed significantly better transfer performance than the interim-math group (M = 35.71, SD = 11.44), $t(39) = 5.72$, $p < .001$, d = 1.83, and the interim-restudy group (M = 49.78, SD = 23.33) also showed significantly better transfer performance than the interim-math group, $t(38) = 2.46$, $p = .019$, $d = 0.80$. The mean difference between the interim-test and interim-restudy group was not significant, $t(37) = 1.30$, $p = .201$. The main effect of the section was also significant, $F(1, 57) = 15.42$, $p < .001$, $\eta_p^2 = .213$, such that overall performance was better on the final test regarding the materials of Section B (M = 52.64, SD = 24.52) than those of Section A (M = 42.29, SD = 18.82), regardless of interim activity. More interestingly, there was a significant interaction effect between interim activity and section, $F(2, 57) = 8.26$, $p = .001$, η_p^2 = .225, implying that the effect of interim activity differed depending on the section.

Different results were obtained for the problems of Sections A and B. For Section A, the participants who were administered an interim-test (M = 45.21, SD = 16.29) solved significantly more transfer problems correctly than those in the interim-math condition (M = 32.54, SD = 12.33), $t(39) = 2.82$, $p = .008$, $d = 0.90$. Likewise, the participants who restudied the earlier materials (M = 50.00, SD = 22.99) solved significantly more transfer problems correctly than those in the interim-math condition, $t(38) = 3.03$, $p = .004$, $d = 0.98$. However, the mean difference between the interim-test and interim-restudy group was not significant, $t(37)$ = 0.75, $p = .456$.

On the other hand, for Section B, the participants who were given an interim-test (M = 70.00, SD = 17.86) solved significantly more transfer problems correctly than those in the interim-restudy condition (M = 49.56, SD = 26.46), $t(37) = 2.84$, p = .007, d = 0.93, and those in the interim-math condition (M = 38.89, SD = 24.52), $t(39) = 5.52$, $p < .001$, $d = 1.77$. The mean difference between the two latter groups was not significant, $t(38) = 1.49$, $p = .142$.

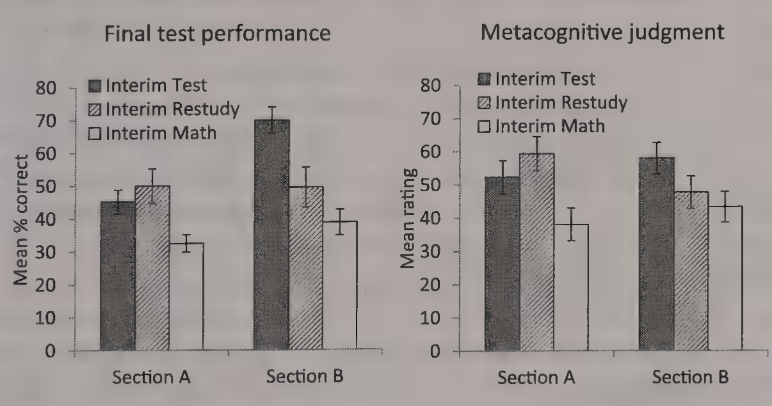

Figure 4. Mean percentage of the correct responses in the final transfer test on Sections A and B (left) and mean ratings of the metacognitive judgments (right) in the interim-test, interim-restudy, and interim-math conditions of Experiment 3. Error bars represent 1 *SEM*.

Discussion

Experiment 3 examined the effect of interim activity on the transfer performance of both Sections A and B through a multiple-choice test format. Compared with Experiment 2, the overall performance increased in the final test. Because the interim-math condition was identical in both Experiments 2 and 3, the increased performance in Experiment 3 can be explained by the change in the format of the final test. More specifically, the multiple-choice test allowed us to measure the participants' classifications, which in turn increased the proportion of the correct responses.

More important, in consonance with the findings from Experiments 1 and 2, Experiment 3 demonstrated that interim testing facilitates subsequent learning of new categories. Regarding transfer performance in Section A, both the interim-test and interim-restudy conditions exhibited better performance than the interim-math condition. The observed performance difference was probably because the test and restudy conditions were exposed to the learning materials of Section A twice, whereas the interim-math condition was exposed to them only once. The former two groups had studied longer; thus, exhibiting better transfer performance. However, the interim-test condition did not demonstrate better performance than the interim-restudy condition regarding Section A, suggesting that there was no apparent testing effect. According to the typical testing effect, the tested group usually shows better retention with respect to the tested materials than the restudy group. However, in this experiment, they performed similarly well. One of the possible explanations for this is that the

participants benefited from interleaved study (Kornell & Bjork, 2008), even when there was no test. In this study, the paintings of different artists were intermixed, rather than grouped by artists. Interleaving could have fostered discrimination learning (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012) and/or memory reloading (Bjork & Bjork, 2011). If interleaving was sufficient to facilitate memory reloading in the restudy condition, then it could have created effects similar to the retrieval practice that the test group had for the materials of Section A.

With respect to transfer performance on Section B, different patterns of results were obtained from those of Section A. For Section B, the participants who were given an interim-test on Section A demonstrated better transfer performance than those in both the interim-restudy and interim-math conditions. The latter two groups of participants showed similarly worse performance than the interim-test group. This finding is especially interesting because the performance of the interim-restudy condition was as good as that of the interim-test condition on Section A. Even though they did similarly well on Section A, only the interim-test group exhibited better performance on Section B. The results suggest that, even if the participants learned the earlier materials effectively (as demonstrated in the good transfer performance of Section A), if they were not tested on them, then there would be no facilitation of subsequent learning. Hence, the beneficial effect of interim testing appears to be more likely because of the testing itself, rather than the better encoding of the previously studied materials.

Experiment 4

Experiments 1–3 investigated the interim-test effect on transfer performance; that is, in the final test participants were tested on novel exemplars that they had not previously seen during their study. Even though this study expanded the interim-test effect to category learning, in Experiment 4 we decided to reinvestigate it by including both recognition and transfer tests in the final test. In an educational context, students are often tested with both studied and unstudied items. In other words, teachers usually test the students with studied items to see how well the students remember the materials explicitly covered in class and test with unstudied items to see how well they can transfer what they had learned to new cases. Therefore, the inclusion of both memory and transfer tests will create more educationally relevant test situations.

Method

Participants. In total, 60 undergraduate students (41 women, 19 men; Mean age = 21 years) participated in exchange for course credit or $5. Each participant was randomly assigned to one of the three conditions, for a total of 20 in the interim-test condition, 20 in the interim-restudy condition, and 20 in the interim-math condition.

Design, materials, and procedure. As illustrated in Figure 1, the overall procedure of Experiment 4 was identical to that of Experiment 3, except for one change; that is, it included both recognition and transfer items in the final test. Accordingly, before the final test, the participants made four different metacognitive judgments with regard to how well they believed that they would be able to correctly identify the following: (a) the paintings previously seen during Section A; (b) the paintings previously unseen, but created by the artists studied in Section A; (c) the paintings previously seen during Section B; and (d) the paintings previously unseen, but created by the artists studied in Section B. In the final test, there were a total of 96 paintings, which consisted of 48 old paintings that served as recognition items and 48 new paintings that served as transfer items. For the recognition items, we randomly selected four paintings from those presented during the study session for each of the 12 artists (six per section). For the transfer items, the same set of paintings from Experiment 3 was used for a total of 48 paintings that consisted of four new paintings by each of the 12 artists. Although the participants had never seen these paintings during their previous study, they were created by the artists who were studied. During the test, the recognition and transfer items were not separated, and the participants were not informed whether they were previously seen or unseen paintings. The paintings were presented in a fixed random order, and the test trials were self-paced with no feedback. All the other procedures were identical to those of Experiment 3.

Results

Interim activity performance between Sections A and B. Only the participants in the interim-test condition were tested on Section A, and the mean percentage of correct responses was 63.61 (SD = 24.40). Cronbach's α of the interim test was .931. In the interim-math condition, the participants solved simple arithmetic problems, and the mean percentage of correct responses was 93.89 (SD = 9.43).

Final test performance. Figure 5 (top) presents the mean percentage of correct responses regarding the recognition and transfer items of Sections A and B among the interim-test, interim-restudy, and interim-math conditions. Cronbach's α values of the final test were .849 on Section A and .848 on Section B for recognition items, .744 on Section A and .809 on Section B for transfer items, respectively. A $3 \times 2 \times 2$ mixed ANOVA was conducted on the number of correct responses. Interim activity (test vs. restudy vs. math) was included as a between-subjects factor, while section (A vs. B) and test item type (recognition vs. transfer) were included as within-subject factors. In general, the participants performed better for the recognition items than for the transfer items, $F(1, 57) = 64.79$, $p < .001$, $\eta_p^2 = .532$. The accuracy of the recognition items (M = 56.15, SD = 20.54) was significantly higher than the accuracy of the transfer items (M = 47.05, SD = 17.19). However, the results of the ANOVA revealed that a three-way interaction was not statistically significant, $F < 1$; thus, implying that the interaction pattern between interim activity and section did not differ, depending on the test item type. Accordingly, in the following, the data were collapsed over the item type (recognition vs. transfer) and we report the results obtained from the 3 (test vs. restudy vs. math) $\times$ 2 (Section A vs. Section B) mixed ANOVA.

The two-way mixed ANOVA revealed a significant main effect of the interim activity, $F(2, 57) = 11.79$, $p < .001$, $\eta_p^2 = .293$. Regardless of section, the participants who were given an interim-test (M = 64.64, SD = 19.96) showed significantly better transfer performance on the final test than those in the interim-restudy condition (M = 49.43, SD = 15.40), $t(38) = 2.70$, $p = .010$, d =

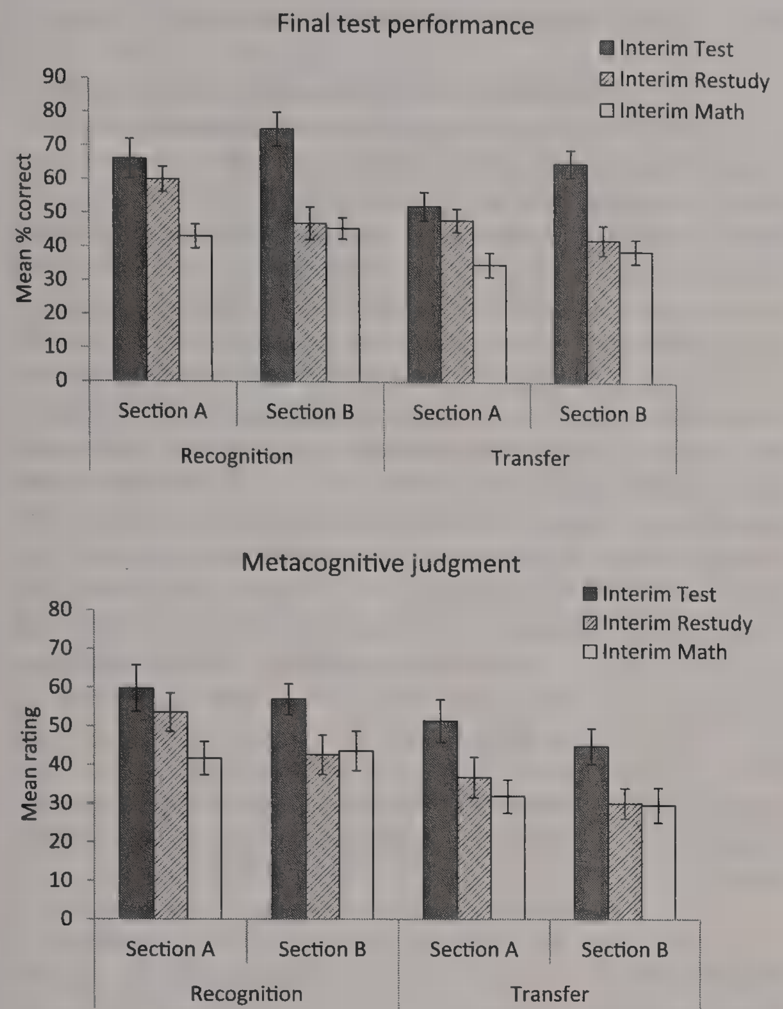

Figure 5. Mean percentage of the correct responses in the final recognition and transfer test on Sections A and B (top) and mean ratings of the metacognitive judgment (bottom) in the interim-test, interim-restudy, and interim-math conditions of Experiment 4. Error bars represent 1 *SEM*.

0.88, and those in the interim-math condition (M = 40.73, SD = 10.48), $t(38) = 4.74$, $p < .001$, $d = 1.54$. Also, the interim-restudy group showed significantly better performance than the interim-math group, $t(38) = 2.09$, $p = .044$, $d = 0.68$. However, the overall effect of the section was not significant, $F < 1$. More interestingly, there was a significant interaction effect between interim activity and section, $F(2, 57) = 7.13$, $p = .002$, $\eta_p^2 = .200$; thus, suggesting that the effect of the interim activity can differ, depending on the section. Indeed, different patterns of results were obtained for Sections A and B.

For Section A, the participants who were given an interim-test (M = 59.17, SD = 21.78) solved significantly more problems correctly than those in the interim-math condition (M = 39.06, SD = 15.23), $t(38) = 3.38$, $p = .002$, $d = 1.09$. Likewise, the participants who restudied the earlier materials (M = 54.06, SD = 14.97) solved significantly more problems correctly than those in the interim-math condition, $t(38) = 3.14$, $p = .003$, $d = 1.02$. However, the mean difference between the interim-test and interim-restudy groups was not significant, $t(38) = 0.87$, $p = .393$.

For Section B, different patterns of results were obtained. The participants who were given an interim-test (M = 70.10, SD = 19.62) solved significantly more problems correctly than those in the interim-restudy condition (M = 44.79, SD = 19.95), $t(38)$ = 4.05, $p < .001$, $d = 1.31$, and those in the interim-math condition (M = 42.40, SD = 14.67), $t(38) = 5.06$, $p < .001$, $d = 1.64$. However, the mean difference between the two latter groups was not significant, $t(38) = 0.43$, $p = .668$.

Discussion

To investigate the interim-test effect on memory and transfer, Experiment 4 included both recognition and transfer items in the final test. The results illustrated that the patterns of results were very similar between the recognition and transfer items. Regarding Section A, the interim-test and interim-restudy conditions exhibited better recognition and transfer than the interim-math condition. Hence, as in Experiment 3, we did not obtain the typical testing effect when comparing the performance between the interim-test and interim-restudy conditions.

However, for Section B, only the interim-test condition (not the restudy condition) illustrated better recognition and transfer performance than the interim-math condition. The results replicated the general findings of Experiments 1–3 in that there is a beneficial effect of an interim test on transfer performance in category learning. The findings were also expanded by indicating the beneficial effects of interim testing on recognition. Hence, the results suggest that interim testing facilitates not only the learning of specific instances but also the generalization of such learning.

Metacognitive Judgments of Experiments 1–4

Experiment 1

Two participants in the interim-test condition did not report their metacognitive judgments. Thus, only the data from 35 participants were included in the data analysis. Figure 2 (right) presents the mean ratings of the metacognitive judgments in the interim-test and interim-math conditions. An independent t test was conducted on the ratings. The participants who were given an interim-test on Section A (M = 47.94, SD = 21.67) made slightly higher predictions for their performance on Section B than those in the interim-math condition (M = 42.81, SD = 25.03). However, the mean difference was not significant, $t(33) = 0.65$, $p = .52$.

The metacognitive results of Experiment 1 showed that the participants in the interim-math condition predicted their transfer performance as good as those in the interim-test condition, although their actual performance was significantly worse than the test condition. In fact, the interim-math group overestimated their competence, whereas the interim-test group underestimated their competence. In this regard, since the interim-math condition did not have an opportunity to test their own learning, they might have experienced the illusion of competence (Koriat & Bjork, 2005). In contrast, although the tested group did not have a test experience with respect to Section B, the interim-test experience on Section A might have alleviated foresight bias (Koriat & Bjork, 2006) and/or decreased overconfidence (Castel, 2008).

Experiment 2

Figure 3 (right) shows the mean ratings of metacognitive judgments for Sections A and B in the interim-test and interim-math conditions. A 2 (test vs. math) × 2 (Section A vs. B) mixed ANOVA was conducted on the ratings. There was only a margin-

ally significant effect of interim activity, $F(1, 28) = 3.05$, $p = .092$, $\eta_p^2 = .098$, in that the participants who were given an interim-test ($M = 40.00$, $SD = 16.13$) predicted their performance to be higher than those who were not administered the test ($M = 29.69$, $SD = 16.13$), regardless of section. There was neither a main effect of section, $F(1, 28) = 2.37$, $p = .135$, nor an interaction effect, $F < 1$; thus, post hoc comparisons were not conducted.

The results showed no significant mean differences among the different conditions or sections. Although the actual transfer performance of the interim-test group was significantly better than that of the interim-math group, the mean differences in their metacognitive judgments did not attain a significant level. This result is consistent with Experiment 1 as the actual performance is not reflected in the metacognitive judgments. One interesting observation is that the interim-test group exhibited significantly better transfer performance on Section B than Section A (as shown on the left side of Figure 3), but their metacognitive judgments were not different between these two sections (as depicted on the right side of Figure 3). This result suggests that the participants were, perhaps, oblivious of the beneficial effects of interim testing on subsequent learning.

Experiment 3

Figure 4 (right) presents the mean ratings of the metacognitive judgments for Sections A and B in the interim-test, interim-restudy, and interim-math conditions. A 3 (test vs. restudy vs. math) × 2 (Section A vs. B) mixed ANOVA was conducted on the ratings. There was a significant main effect of interim activity, $F(2, 57) = 3.36$, $p = .042$, $\eta_p^2 = .105$, but not an effect of section, $F < 1$. Regardless of section, the interim-test group ($M = 55.13$, $SD = 18.50$) predicted their performance to be significantly higher than the interim-math group ($M = 40.67$, $SD = 20.97$), $t(39) = 2.34$, $p = .025$, $d = 0.75$. Also, the interim-restudy group ($M = 53.55$, $SD = 19.19$) predicted their performance higher than the interim-math group, but the mean difference between these two groups was only marginally significant, $t(38) = 2.02$, $p = .05$. The mean difference between the interim-test and interim-restudy group was not also significant, $t(37) = 0.26$, $p = .796$. More interestingly, the interim activity by section interaction was significant, $F(2, 57) = 4.88$, $p = .011$, $\eta_p^2 = .146$; thus, implying that the effect of interim activity on metacognitive judgments differed depending on the section.

Similar to the actual test performance, different patterns of judgments were observed for Sections A and B. For Section A, the participants who were administered an interim-test ($M = 52.25$, $SD = 23.37$) predicted their transfer performance higher than those in the interim-math condition ($M = 38.05$, $SD = 23.97$), but the mean difference was only marginally significant, $t(39) = 1.92$, $p = .062$, $d = 0.61$. In addition, the participants who restudied the earlier materials ($M = 59.36$, $SD = 19.76$) predicted their performance to be significantly higher than those in the interim-math condition, $t(38) = 3.05$, $p = .004$, $d = 0.99$. The mean difference between the interim-test and interim-restudy group was not significant, $t(37) = 1.03$, $p = .312$. In contrast, for Section B, even though the interim-test participants ($M = 58.00$, $SD = 18.17$) predicted their transfer performance on Section B to be higher than the interim-restudy participants ($M = 47.74$, $SD = 22.56$), who predicted their performance higher than those in the interim-math condition ($M = 43.29$, $SD = 23.07$), the mean ratings among the three conditions were not significantly different, $F(2, 57) = 2.53$, $p = .089$.

Similar to Experiments 1 and 2, metacognitive judgments made by the participants did not reflect their actual transfer performance. However, the general patterns of actual performance and metacognitive judgments were surprisingly similar. For Section A, the interim-restudy group predicted their performance to be higher than the interim-math group, and the actual performance of the former group was indeed better than the latter group. Although the interim-test group also made higher predictions, and their actual performance was indeed better than the interim-math group, the mean difference of the predictions was not significant between the two groups. Moreover, for Section B, the interim-test group made higher predictions, and their actual performance was indeed better than the other groups. However, there were no significant differences in terms of the predictions among the three conditions. This null effect of interim activity on metacognitive judgments was mostly based on the fact that the interim-test group underestimated their competence of Section B. Although their actual performance was significantly worse in Section A ($M = 45\%$) than Section B ($M = 70\%$), their predictions were not that different in Section A ($M = 52\%$) and Section B ($M = 58\%$). This finding is consistent with Experiments 1 and 2 as the participants were probably unaware of the beneficial effects of interim testing on subsequent learning.

Experiment 4

Four participants did not report their metacognitive judgments (one in the interim-test condition, one in the interim-study condition, and two in the interim-math condition). As a result, only the data from 56 participants were included in the analyses. Figure 5 (bottom) shows the mean ratings of the metacognitive judgments for the recognition and transfer items of Sections A and B among the interim-test, interim-restudy, and interim-math conditions. A 3 (test vs. restudy vs. math) × 2 (Section A vs. Section B) × 2 (recognition vs. transfer) mixed ANOVA was conducted on the ratings. The ANOVA results revealed that there was a significant main effect of item type, $F(1, 53) = 64.25$, $p < .001$, $\eta_p^2 = .548$, such that the participants gave significantly higher ratings for the recognition items ($M = 50.24$, $SD = 19.70$) than the transfer items ($M = 37.63$, $SD = 20.58$). Similar to the actual test performance, a three-way interaction was not statistically significant, $F(2, 53) = 1.19$, $p = .310$; thus, implying that the patterns of results did not differ, depending on the item type. Accordingly, in the following analyses, the data were collapsed over the item type (recognition vs. transfer), and we report the results obtained from the 3 (test vs. restudy vs. math) × 2 (Section A vs. Section B) mixed ANOVA.

The two-way mixed ANOVA revealed a significant main effect of interim activity, $F(2, 53) = 4.54$, $p = .015$, $\eta_p^2 = .146$. This main effect was because of higher predictions made by the interim-test group ($M = 53.89$, $SD = 19.26$). The interim-test group made significantly higher predictions than the interim-restudy group ($M = 40.79$, $SD = 18.56$), $t(36) = 2.13$, $p = .04$, $d = 0.71$, and the interim-math group ($M = 36.76$, $SD = 16.41$), $t(35) = 2.90$, $p = .006$, $d = 0.98$. Moreover, there was neither a main effect of section, $F(1, 53) = 3.63$, $p = .062$, nor an interaction effect between the interim activity and section, $F < 1$. Because the

interaction effect was not statistically reliable, post hoc comparisons were not conducted.

As observed in Experiment 3, metacognitive judgments made by the participants did not exactly reflect their actual performance. However, the general patterns were surprisingly similar between the actual performance and metacognitive judgments. Although statistical tests failed to show significant interactions between section and interim activity, the rank order of ratings among the three conditions were very similar to that of actual performance. The interim-test condition always showed the best performance, and it always provided the highest metacognitive judgments compared with the other conditions. The greatest difference between the actual test performance and metacognitive judgments was that the interim activity by section interaction was significant in actual test performance, whereas it was not significant in metacognitive judgments. This was because, in the actual performance, the benefit of interim testing was only apparent in Section B (not in Section A) compared with the restudy condition, whereas in metacognitive judgments, it was less apparent.

Discussion

Across the four experiments, the metacognitive measures provided by the participants did not reflect their actual performance, which suggests that the participants were unaware about the beneficial effects of interim testing on their subsequent learning. In Experiments 1 and 2, we observed that the participants in the interim-math condition predicted their performance regarding the recently studied categories of the final target section (i.e., Section B) as high as the participants who were given an interim-test on the preceding section (i.e., Section A). Because both groups of participants had an equal amount of study time and exposure to the final target section, it may have been reasonable for them to predict similar levels of performance. However, the actual performance was much better in the tested group; thus, implying that these participants were unaware about the interim-test effect on subsequent learning. Experiments 3 and 4 also demonstrated similar patterns of metacognitive judgments in that the test group was unaware about the beneficial effects of interim testing on subsequent new learning. In addition, the participants in the interim-test condition showed significantly better transfer performance on Section B, but their metacognitive judgments on Section B were not always significantly higher than the other conditions. Such unawareness about the beneficial effects of interim testing, however, may be simply because of the design of the current study. The current study always adopted a between-subjects design and participants never knew about the comparison condition. While not knowing the other conditions, participants may tend to shift toward using the middle of rating scales and such tendency might have created a null effect in the current study. Yan, Bjork, and Bjork (2016, Experiment 6) also reported that even in a within-subject design metacognitive experiences could be influenced by the order of conditions participants were exposed to.

One interesting observation was that there was a high level of similarity in terms of the rank order of the interim-test, interim-restudy, and interim-math conditions between actual performance and metacognitive judgments, although the statistical significance of their mean differences did not match. In both Experiments 3 and 4, the interim-test group demonstrated higher ratings than the interim-restudy and interim-math groups on almost every type of metacognitive judgment (with one exception of Section A in Experiment 3). The finding that the test group made higher predictions than the restudy group for Section B can be viewed as contradictory to previous research. People who restudy materials often experience the illusion of competence (e.g., Yue et al., 2015), and they are generally unaware about the benefits of the testing effect (for a review, see Karpicke et al., 2009). In the present study, the interim-restudy group actually performed as well as the interim-test group on the restudied items. Thus, they did not appear to experience the illusion of competence.

The results also suggest that metacognition in category learning can differ from metacognition of learning that mostly involves materials, such as word lists and short text passages, which were typically used in previous investigations of metacognition (Dunlosky & Metcalfe, 2009). In recent studies, metacognitive judgments have been examined at the level of categories using materials of bird families (e.g., Jacoby et al., 2010; Tauber & Dunlosky, 2015; Wahlheim, Dunlosky, & Jacoby, 2011). For example, Jacoby et al. (2010) investigated metacognition by examining the participants' predictions of their ability to identify novel exemplars from the studied categories, after which the results showed that the participants were aware about the beneficial effects of testing. The participants seemed to be aware of the difficulty differences at classifying exemplars across categories. However, one big difference between the Jacoby et al.'s and current study is that the former involved having a participant judge his or her learning of each category (such measure is called category learning judgment: CLJ), whereas the latter involved having participants make a global judgment of learning. Future studies need to investigate metacognition with more diverse types of materials and methods of measuring metacognitive judgments for expanding our understanding of metacognition.

General Discussion

The four experiments in this study demonstrated that the interim testing of prior categories facilitates the learning of subsequently presented new categories. The results extend the findings of previous studies on the interim-test effect (e.g., Szpunar et al., 2008; Wissman et al., 2011) by indicating that the beneficial effects of interim testing can be generalized into category learning. However, the metacognitive measures provided by the participants did not reflect their actual performance, suggesting that the participants were unaware about the beneficial effects of interim testing.

For investigating category learning, the current study used a painting-style learning task wherein the participants had to first learn the painting styles of different artists by studying specific painting examples across two separate sections (Sections A and B); subsequently, this learning was applied to other new instances of the studied categories. Here, different artists served as the different categories, while the specific paintings served as the exemplars of the categories. Depending on whether the participants were tested or not tested on the categories in Section A before moving on to study the new categories in Section B, the participants demonstrated different levels of transfer performance on the final transfer test regarding Section B. The results of Experiment 1 revealed that the participants who were administered an interim-test on Section A showed better transfer performance

on Section B than the participants who were not tested. Experiment 2 replicated this finding when the final transfer test included all the learned categories from both Section A and Section B, while Experiments 3 and 4 replicated the result when the final test format was changed to a multiple-choice test. Experiments 3 and 4 also established that the beneficial effect of interim testing was because of the testing itself, rather than the high level of initial learning by comparing the interim-test and interim-restudy conditions. More specifically, the participants who had interim-restudy regarding Section A transferred as well as the participants who had interim-test regarding Section A, when they were administered a transfer test on the categories of Section A. However, their transfer performance on Section B was much worse than the interim-test group of participants. Experiment 4 further demonstrated the same patterns of results when we examined recognition as well as transfer performance. As observed in several previous studies on the interim-test effect (e.g., Szpunar et al., 2008; Wissman et al., 2011; Yue et al., 2015), interim testing benefited the retention of materials that were studied after the interim test. Altogether, the results from the four experiments suggest that interim testing facilitates not only subsequent learning of specific instances but also the transfer of such learning.

The overarching goal of the present research was to explore whether the interim-test effect would extend into category learning. Despite the fact that an investigation of the underlying mechanisms goes beyond the scope of the current study, it is important to discuss what might have caused the beneficial effects of interim testing on subsequent learning of new materials by examining the learning conditions in this study. There could be several possible hypotheses on what causes the interim-test effect. First, one may suspect that interim testing increases test expectancy. In this study, across the four experiments, the participants were forewarned that they would be tested after completion of the study sessions. Because it is believed that all the participants were aware about the upcoming test, the test expectancy explanation seems unlikely. Second, interim testing might have worked as an intervening activity that separates the two sections, which in turn could reduce the interference effects between the sections. To address this possibility, the present research design always included a control condition in which the participants were administered an interim-math activity instead of an interim test. However, because the interim-math group always performed worse than the test group in the final test, the intervening activity itself is less likely to explain the interim-test effect. Third, interim testing might have provided an additional exposure to the studied materials, which in turn could have promoted better encoding of such materials. Because the participants in the test condition were exposed to the studied materials twice (once during their study and once during the interim test), they had a higher level of exposure to the learning materials than the participants in the interim-math condition. Meanwhile, the interim-math condition participants were exposed to the studied materials only once during their study since they were not tested. To address this unfair advantage of the interim-test group, Experiments 3 and 4 included an interim-restudy condition, in addition to the interim-math condition, as comparison groups. The results revealed that the interim-restudy group showed better transfer performance only on the previously restudied categories (not on the subsequently studied new categories) than those in the interim-math condition. Accordingly, it was concluded that the interim testing effect was more likely to be because of the testing itself, rather than because of the better encoding of the initial learning.

A more plausible explanation for the enhanced learning after interim testing involves metacognitive benefits. In general, testing is known to improve metacognitive knowledge (Karpicke, 2009). While being tested on the previously studied materials, the participants may be able to evaluate their earlier learning strategies and adjust them, if they believe that they were not good. Pyc and Rawson (2010, 2012) proposed a mediator-shift hypothesis, which states that retrieval failure during practice encourages individuals to shift from less effective mediators to more effective ones, when given a restudy opportunity. Consistent with this hypothesis, Soderstrom and Bjork (2014) reported that, after receiving a review test, the participants switched to more effective encoding strategies. In the present study, only the tested group perhaps had an opportunity to evaluate their own strategies, which might have affected their subsequent learning strategy, when given the new learning materials.

Moreover, the participants might have realized that the test was not as easy as they had expected. As a result, they might have decided to put more encoding effort into their subsequent learning. Previous studies have shown that students can be less confident in their learning after testing (Finn & Metcalfe, 2007, 2008; Koriat & Bjork, 2006; Koriat et al., 2002; Meeter & Nelson, 2003). In addition, difficulties encountered during the intervening test might have prepared the students to learn better by affecting their subsequent study strategies; that is, interim testing might have served as a preparation for future learning (see Bransford & Schwartz, 1999). Kapur (2008, 2011) also proposed that failure experience during the initial learning phases can encourage students to learn better in subsequent learning phases by helping them better attend to critical features of the to-be-learned concept. This phenomenon is what Kapur termed as *productive failure*. Although the experience of failure discussed in his research refers to the failure of generating valid, problem-solving methods during the invention phase (similar to discovery learning), the test situation in the present study could have also caused the participants to experience failure, especially if they believed that the test was more difficult than expected. In the current study, because the interim test was always a cued-recall test, most of the participants found the task to be quite challenging, as shown in their poor performance of the interim tests in the four experiments. Future studies should examine what mechanisms underlie the interim-test effects and how interim tests can influence subsequent encoding strategies.

One remaining issue to address is how interim testing affected different components of the task used in the current study. The painting-style learning task consists of largely three components: learning the content of the category (i.e., the artist' styles), learning the category names (i.e., the artists' names), and finally learning the association between a given category and a category's name (i.e., linking the style and artist's name). Interim testing could have affected some or all of these components, but the current study did not separate the components involved in the task to examine which one was affected by interim testing. One possibility is that participants in the interim-test condition learned the artist names better than the other conditions. However, Yan et al.'s study (2016, Experiment 5) showed even when participants had a preliminary name-learning phase, superiority of interleaved schedule remained

over blocked schedule in inductive learning of painting styles. Another possibility is that participants in the interim-test, interim-restudy, and interim-math conditions might have learned the artist names and artist styles themselves to a similar degree, but that the link between the name and the style might have been stronger in the interim-test condition than the other conditions. If it is the case, a preliminary name-learning phase would not remove the observed interim-test effect. Future studies will need to investigate which components of the task are affected by interim testing.

Limitations and Future Directions

One limitation of this study is that the type of interim test was always a cued-recall test. The recall test was chosen based on the procedures observed in previous studies on the interim-test effect (e.g., Cho et al., 2017; Wissman et al., 2011). Although cued-recall tests are generally known to be more effective learning events than other forms of testing (e.g., Bjork & Whitten, 1974; Carpenter, Pashler, & Vul, 2006; Glover, 1989; Rowland, 2014), different types of testing could also be effective (or even more effective) than cued-recall tests. For example, Little and Bjork (2016) found that multiple-choice tests were more effective than cued-recall tests when multiple-choice questions involved competitive and related alternatives. According to the transfer-appropriate processing view (Morris, Bransford, & Franks, 1977), the effects of testing may be dependent on the degree to which how the cues given on an interim test correspond to those given on a final test. On the other hand, Carpenter and Delosh (2006) emphasized the importance of elaborative processing for testing effects to occur by demonstrating that the provision of impoverished cues during intervening tests can enhance subsequent retention. Furthermore, the type of practice tests may give students a different expectancy about an upcoming test, which in turn can influence a change in their encoding strategies (Finley & Benjamin, 2012; Storm, Hickman, & Bjork, 2016). To obtain optimal forms of interim tests, future research should include diverse forms of interim tests and test their effects on subsequent learning.

Another limitation of this study is that the final tests were administered almost immediately after the last study section. The interval between the last study section and the final test was as long as the duration that the participants took to provide their metacognitive judgments. Although we believe interim tests could be used in real educational settings to encourage students to be engaged in more effective study strategies, the ultimate goal of education is not a short-term enhancement, but long-term enhanced learning. Hence, future research should investigate how long the benefits of interim tests last on retention and transfer in category learning.

Conclusion

Overall, the results of this study support the idea that tests are powerful tools for learning. Testing enhances not only the learning of tested items (as shown in extensive research on the typical testing effect) but also the learning of untested items that are subsequently presented (as shown in research on the interim-test effect). The current study is, to the best of our knowledge, the first to demonstrate that interim testing on previously studied categories can enhance the subsequent study of new categories. Therefore, the findings extend the interim-test effect into category learning by showing that the beneficial effects of interim testing not only occur for the learning of specific instances but also for the generalization of such learning. Interim testing appears to help students to learn better and such a preparation is not obtained from interim restudy. From an educational perspective, this study suggests that educators may want to use tests as a preparation for subsequent learning. For instance, to enhance student learning, instructors can divide a class into smaller units and administer interim tests on the preceding units before studying subsequent units.

References

Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., III, & McDermott, K. B. (2008). Examining the testing effect with open-and closed-book tests. *Applied Cognitive Psychology, 22,* 861–876. http://dx.doi.org/10.1002/acp.1391

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 940–945. http://dx.doi.org/10.1037/a0029199

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30,* 666–677. http://dx.doi.org/10.3758/BF03196423

Baars, M., Van Gog, T., De Bruin, A., & Paas, F. (2014). Effects of problem solving after worked example study on primary school children's monitoring accuracy. *Applied Cognitive Psychology, 28,* 382–391. http://dx.doi.org/10.1002/acp.3008

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition, 41,* 392–402. http://dx.doi.org/10.3758/s13421-012-0272-7

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York, NY: Worth Publishers.

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123–144). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Whitten, W. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189. http://dx.doi.org/10.1016/0010-0285(74)90009-7

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24,* 61–100.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1118–1133. http://dx.doi.org/10.1037/a0019902

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1547–1552. http://dx.doi.org/10.1037/a0024140

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition, 34,* 268–276. http://dx.doi.org/10.3758/BF03193405

Carpenter, S. K., Pashler, H., & Vul, E. (2006). What types of learning are enhanced by a cued recall test? *Psychonomic Bulletin & Review, 13,* 826–830. http://dx.doi.org/10.3758/BF03194004

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review, 22,* 281–288. http://dx.doi.org/10.3758/s13423-014-0676-4

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition, 36,* 429–437. http://dx.doi.org/10.3758/MC.36.2.429

Cho, K. W., Neely, J. H., Crocco, S., & Vitrano, D. (2017). Testing enhances both encoding and retrieval for both tested and untested items. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 70,* 1211–1235. http://dx.doi.org/10.1080/17470218.2016.1175485

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition.* Thousand Oaks, CA: Sage.

Finley, J. R., & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 632–652. http://dx.doi.org/10.1037/a0026215

Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 238–244. http://dx.doi.org/10.1037/0278-7393.33.1.238

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58,* 19–34. http://dx.doi.org/10.1016/j.jml.2007.03.006

Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81,* 392–399. http://dx.doi.org/10.1037/0022-0663.81.3.392

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 801–812. http://dx.doi.org/10.1037/a0023219

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 290–296. http://dx.doi.org/10.1037/a0028468

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18,* 879–919. http://dx.doi.org/10.2466/pr0.1966.18.3.879

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1441–1451. http://dx.doi.org/10.1037/a0020636

Kang, S. H., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19,* 528–558. http://dx.doi.org/10.1080/09541440601056620

Kang, S. H., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97–103. http://dx.doi.org/10.1002/acp.1801

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26,* 379–424. http://dx.doi.org/10.1080/07370000802212669

Kapur, M. (2011). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39,* 561–579. http://dx.doi.org/10.1007/s11251-010-9144-3

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469–486. http://dx.doi.org/10.1037/a0017341

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17,* 471–479. http://dx.doi.org/10.1080/09658210802647009

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93,* 329–343. http://dx.doi.org/10.2307/1422236

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 187–194. http://dx.doi.org/10.1037/0278-7393.31.2.187

Koriat, A., & Bjork, R. A. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1133–1145. http://dx.doi.org/10.1037/0278-7393.32.5.1133

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131,* 147–162. http://dx.doi.org/10.1037/0096-3445.131.2.147

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19,* 585–592. http://dx.doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging, 25,* 498–503. http://dx.doi.org/10.1037/a0017807

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989–998. http://dx.doi.org/10.1037/a0015729

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17,* 493–501. http://dx.doi.org/10.1080/09658210902832915

Little, J. L., & Bjork, E. L. (2016). Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition, 44,* 1085–1101. http://dx.doi.org/10.3758/s13421-016-0621-z

Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition, 43,* 85–98. http://dx.doi.org/10.3758/s13421-014-0453-7

McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16,* 192–201. http://dx.doi.org/10.1016/0361-476X(91)90037-L

Mcdaniel, M. A., Roediger, H. L., III, & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14,* 200–206. http://dx.doi.org/10.3758/BF03194052

Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologica, 113,* 123–132. http://dx.doi.org/10.1016/S0001-6918(03)00023-4

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519–533. http://dx.doi.org/10.1016/S0022-5371(77)80016-9

Murphy, G. (2004). *The big book of concepts.* Cambridge, MA: MIT Press.

Nunes, L. D., & Weinstein, Y. (2012). Testing improves true recall and protects against the build-up of proactive interference without increasing false recall. *Memory, 20,* 138–154. http://dx.doi.org/10.1080/09658211.2011.648198

Pastötter, B., & Bäuml, K. H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5,* 286. http://dx.doi.org/10.3389/fpsyg.2014.00286

Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K. H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 287–297. http://dx.doi.org/10.1037/a0021801

Pastötter, B., Weber, J., & Bäuml, K. H. T. (2013). Using testing to improve learning after severe traumatic brain injury. *Neuropsychology, 27,* 280–285. http://dx.doi.org/10.1037/a0031797

Putnam, A. L., & Roediger, H. L., III. (2013). Does response mode affect amount recalled or the magnitude of the testing effect? *Memory & Cognition, 41,* 36–48. http://dx.doi.org/10.3758/s13421-012-0245-x

Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science, 330,* 335. http://dx.doi.org/10.1126/science.1191465

Pyc, M. A., & Rawson, K. A. (2012). Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 737–746. http://dx.doi.org/10.1037/a0026166

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review, 24,* 419–435. http://dx.doi.org/10.1007/s10648-012-9203-1

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied, 15,* 243–257. http://dx.doi.org/10.1037/a0016496

Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15,* 20–27. http://dx.doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III, & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249–255. http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432–1463. http://dx.doi.org/10.1037/a0037559

Rowland, C. A., Littrell-Baez, M. K., Sensenig, A. E., & DeLosh, E. L. (2014). Testing effects in mixed- versus pure-list designs. *Memory & Cognition, 42,* 912–921. http://dx.doi.org/10.3758/s13421-014-0404-3

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73,* 99–115. http://dx.doi.org/10.1016/j.jml.2014.03.003

Storm, B. C., Hickman, M. L., & Bjork, E. L. (2016). Improving encoding strategies as a function of test knowledge and experience. *Memory & Cognition, 44,* 660–670. http://dx.doi.org/10.3758/s13421-016-0588-9

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 6313–6317. http://dx.doi.org/10.1073/pnas.1221764110

Szpunar, K. K., McDermott, K. B., & Roediger, H. L., III. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392–1399. http://dx.doi.org/10.1037/a0013082

Tauber, S. K., & Dunlosky, J. (2015). Monitoring of learning at the category level when learning a natural concept: Will task experience improve its resolution? *Acta Psychologica, 155,* 8–18. http://dx.doi.org/10.1016/j.actpsy.2014.11.011

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & Cognition, 43,* 27–38. http://dx.doi.org/10.3758/s13421-014-0455-5

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39,* 750–763. http://dx.doi.org/10.3758/s13421-010-0063-y

Weinstein, Y., Gilmore, A. W., Szpunar, K. K., & McDermott, K. B. (2014). The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1039–1048. http://dx.doi.org/10.1037/a0036164

Weinstein, Y., McDermott, K. B., & Szpunar, K. K. (2011). Testing protects against proactive interference in face-name learning. *Psychonomic Bulletin & Review, 18,* 518–523. http://dx.doi.org/10.3758/s13423-011-0085-x

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review, 18,* 1140–1147. http://dx.doi.org/10.3758/s13423-011-0140-7

Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145,* 918–933. http://dx.doi.org/10.1037/xge0000177

Yue, C. L., Soderstrom, N. C., & Bjork, E. L. (2015). Partial testing can potentiate learning of tested and untested material from multimedia lessons. *Journal of Educational Psychology, 107,* 991–1005. http://dx.doi.org/10.1037/edu0000031

Received January 12, 2017
Revision received April 21, 2017
Accepted April 21, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 218–232

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000207

Anthropomorphism in Decorative Pictures: Benefit or Harm for Learning?

Sascha Schneider, Steve Nebel, Maik Beege, and Günter Daniel Rey
Chemnitz University of Technology

When people attribute human characteristics to nonhuman objects they are amenable to anthropomorphism. For example, human faces or the insertion of personalized labels are found to trigger anthropomorphism. Two studies examine the effects of these features when included in decorative pictures in multimedia learning materials. In a first experiment, 81 university students were randomly assigned to 1 cell of a 2 (human faces vs. no faces in pictures) × 2 (personalized vs. nonpersonalized labels of pictures) between-subjects, factorial design. In addition to learning performance, cognitive, motivational, and emotional impacts of anthropomorphism are examined. Results show that both human faces and anthropomorphic labels were able to increase the learning performance on cognitive assessments. However, only human faces were able to influence motivational and emotional ratings significantly. In a second experiment, 108 secondary school students were randomly assigned to 3 groups (anthropomorphized pictures, nonanthropomorphized pictures, and no pictures) in order to evaluate possible advantages of anthropomorphism in decorative pictures in learning materials. Results show again that anthropomorphized pictures are better for learning than nonanthropomorphized pictures and also better than a control group. Results are discussed in the light of a debate on the inclusion or exclusion of decorative pictures.

Educational Impact and Implications Statement

This research reveals that incorporating decorative pictures within multimedia materials is beneficial for learning when tendencies of attributing human characteristics are triggered through specific picture features. Both integrated human faces and personalized labels are found to enhance learning performance and improve learners' affect and motivation in contrast to pictures without these features or materials without decorative pictures in 2 experiments. In conclusion, decorative pictures may be used in order to make learning materials more appealing if boundary conditions like the degree of anthropomorphism were taken into account.

Keywords: anthropomorphism, personalization, decorative pictures, seductive detail effect, multimedia learning

Within many learning materials textual information is illustrated by instructional or decorative pictures. According to Takahashi (1995), instructional pictures and decorative pictures must be distinguished because of a difference in their main function: providing information versus enabling an aesthetic experience. However, Lenzner, Schnotz, and Müller (2013) suggest instead to use both main functions (information provision and decoration) as two orthogonal dimensions, which do not exclude each other. In conclusion, only purely decorative pictures can be defined as pictures which do not provide information (or at least no learning-relevant information), but are included to enrich learning materials with pictures. The majority of studies which examined decorative pictures revealed learning-inhibiting effects. These studies are often theoretically based on the *seductive detail effect* (Harp & Mayer, 1998; for a metaanalytical overview see Rey, 2012), which states that learning-irrelevant but interesting elements are detrimental for learning. However, decorative picture studies focused little on the moderating role of different picture design features: for example, the amount of displayed humans or emotional effects (Schneider, Nebel, & Rey, 2016). The implementation of anthropomorphic features in learning-relevant pictures, for example, is shown to enhance learning performance through attributing characteristics of humans to the learning material (e.g., Mayer & Estrella, 2014; Park, Knörzer, Plass, & Brünken, 2015). In the present study, two design features of anthropomorphism (i.e., human shapes and personalization) were implemented in decorative pictures in order to examine the effects of these learning-enhancing features in decorative pictures.

This article was published Online First April 27, 2017.

Sascha Schneider, Steve Nebel, Maik Beege, and Günter Daniel Rey, E-Learning and New Media, Faculty of Humanities, Chemnitz University of Technology.

Correspondence concerning this article should be addressed to Sascha Schneider, E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Straße der Nationen 12, 09111 Chemnitz, Germany. E-mail: sascha.schneider@phil.tu-chemnitz.de

Theories of Learning With Text and Pictures

There is a long tradition within the learning sciences to examine effects of picture and text combinations (e.g., Samuels, 1970; Schüler, Arndt, & Scheiter, 2015). The integration of text and

pictures within learning materials is not only a central issue for school textbooks but also started new research fields like *multimedia learning*, which is defined as "learning from words and pictures" (Mayer, 2001). Within this field, learning theories, such as the cognitive theory of multimedia learning (CTML; Mayer, 2014a) or the integrated model of text and picture comprehension (Schnotz, 2005), are concerned with the effect of text-picture combinations.

CTML is based on the SOI-assumption, which describes that learners are required to select, organize, and integrate information in order to build and maintain a coherent mental model. In more detail, it has been argued that visual information (i.e., images) and auditory information is encoded via two separate channels which are limited in their capacities. Both channels select information through attentional processes and organize this information to coherent mental models. Both models are then integrated into long term memory by combining this new information with prior knowledge. During the learning process, three cognitive processes, referring to the research of Sweller (e.g., Sweller, 1994, 2010), can be distinguished (Kalyuga, 2011): (a) essential processing, defined as the processing of learning-relevant information (also referred to as intrinsic cognitive load); (b) extraneous processing, caused by suboptimal design of materials (also referred to as extraneous cognitive load); and (c) generative processing, subsumed as all processes which make sense of the essential material (also referred to as germane cognitive load). However, this theory does not include influences of learning-relevant variables like emotion or motivation. As a consequence, researchers like Moreno (2006; cognitive-affective theory of learning with media) and Plass and Kaplan (2015; integrated cognitive affective model of learning with multimedia [ICALM]), suggested theoretical extensions of the CTML. Plass' theory, for example, includes possible influences of emotionally charged learning materials (elicited by, e.g., decorative elements), which might lead to differences in interest or motivation. In the case of effects of decorative pictures, both theories (CTML and ICALM) can substantiate findings as a framework for cognitive and affective learning processes.

Decorative Pictures in Multimedia Learning

Based on Carney and Levin (2002), pictures serve different functions in learning processes according to their contribution of relevant information. While representational, organizational, interpretational, and transformational pictures are directly linked to learning-enhancing processes, decorative pictures are seen as nonrelevant in order to achieve a higher level of knowledge. This separation is based on mainly cognitive aspects of learning, however, other classifications which consider the amount of attention attraction (Levie & Lentz, 1982), or the degree of emotional impact or metacognitive support (Chen & Latham, 2014), are possible. In addition, Takahashi (1995) suggested instead to separate pictures in learning environments between instructional pictures and decorative pictures. While the main function of instructional pictures is provision of information, the main function of decorative pictures is production of aesthetic experiences. This duality was also used by Schneider et al. (2016), who additionally distinguished between positive (conducive) decorative pictures and negative (seductive) decorative pictures, since the duality between positive or negative impacts of decorative pictures is especially mirrored within multimedia research.

For some multimedia researchers, decorative pictures are considered as learning impediments and therefore called seductive (decorative) pictures (e.g., Harp & Mayer, 1998). Although a meta-analysis by Rey (2012) showed that retention and transfer results are impaired by these pictures, it is still not clear what makes pictures more or less seductive or how various influences (e.g., emotional states, motivation) might moderate this effect. In a study by Magner, Schwonke, Aleven, Popescu, and Renkl (2014), geometry principles were taught in two conditions: with and without decorative pictures. Results indicate that only learners with low prior knowledge of the learning content are susceptible to the attention-distraction effects of seductive pictures. However, in this study relevant learning content was implemented into decorative pictures. Moreover, it could be demonstrated that these pictures induce higher situational interest ratings which affects transfer performance positively. This result is consistent with a finding of Park, Kim, Lee, Son, and Lee (2005), where seductive pictures fostered the level of interest without an impediment of learning performance. In conclusion, the impact of decorative pictures is not clear-cut. Some of the studies might have increased the seductive-detail effect through the implementation of other features into nonrelevant pictures, such as anthropomorphic features (e.g., Sung & Mayer, 2012), by drawing more attention toward the included seductive pictures. In some cases, decorative pictures are found to rather support learning through an increase of students' mood, calmness, or alertness (Lenzner et al., 2013); an improvement of attractiveness and situational interest (Male, 2007; Rubens, 2000); or a stimulation of visual aesthetics (Chiaverina, Scott, & Steele, 1997). In addition, decorative pictures are shown to increase learning when instructions are personalized rather than impersonal (Wang & Crooks, 2015). However, none of the studies examined the effects of anthropomorphic features in decorative pictures on accompanied learning information.

Anthropomorphism and Learning

Anthropomorphism is defined as the attribution of uniquely human mental characteristics to nonhumans (Waytz, Klein, & Epley, 2013). These uniquely characteristics are mental states that imply agency, such as intentions, beliefs, or conscious experiences. Examples for this might be feelings like joy or shame (e.g., Farah & Heberlein, 2007). However, triggers are needed to activate the attribution of anthropomorphism (Gilbert & Hixon, 1991; Kim & Sundar, 2012; Waytz et al., 2013). These triggers might stem from the recipient, such as the motivation to understand the behavior of other entities, as well as the characteristics of the entity that is perceived, such as similarities in outer appearances or behavioral components (Waytz, Gray, Epley, & Wegner, 2010). For example, avatars or robots that look like human beings are perceived as more competent and intelligent than avatars without human appearances (Nowak, Hamilton, & Hammond, 2009; Złotowski, Proudfoot, Yogeeswaran, & Bartneck, 2015). Instructions of programs presented with a human voice are also more likely to comply than computer voices (Lee, 2010). Especially human faces or parts of it are shown to trigger anthropomorphism (Gong, 2008). The inclusion of these features is also referred to as the embodiment principle (Mayer, 2014b). In a study by Mayer

and Estrella (2014) students had to learn how viruses attack host cells in bodies. The inclusion of expressive eyes into visualizations of host cells and viruses fostered learning performance and mental effort in contrast to a nonanthropomorphized condition, however, ratings on the appeal or enjoyment of the lessons were not affected. Results by Um, Plass, Hayward, and Homer (2012) indicate that the inclusion of faces (called shapes) into learning materials on how immunization works were able to enhance retention and transfer scores via an enhancement of positive emotions in contrast to a nonface condition. Similar results can be retrieved from two experiments by Plass, Heidig, Hayward, Homer, and Um (2014), who used the same learning materials as the previous mentioned study. While learning was fostered through facelike features, perceived difficulties of the learning materials decreased, and motivation increased for students with anthropomorphic features. Faces, however, did only foster knowledge transfer when no color was added. Park et al. (2015) evaluated this learning material in an eye-tracking study and showed that anthropomorphisms capture learners' attention. This might motivate learners to spend more time with a learning material. In addition, students with face features in their learning materials reported less perceived task difficulty and a higher amount of intrinsic motivation. Sherman and Haidt (2011) point out in their review on the humanizing effects of emotion that positive emotions rather than neutral states evoke anthropomorphism. Haaranen, Ihantola, Sorva, and Vihavainen (2015) suggested to pay attention to learners' study time, since anthropomorphic graphics reduced time on task in their study in contrast to a group without graphics, while the learners' comprehension scores were not affected. In conclusion, the authors suggest controlling for time on task.

Personalization as a Form of Anthropomorphism

The personalization principle, defined as the method of including conversational instead of formal text cues, such as "my" and "your" instead of articles, is shown to enhance learners' performance (for a meta-analytic overview, see Ginns, Martin, & Marsh, 2013). These features make an addressing of learners and the personality of the instructor more salient. Because of this familiarity to other social (human) situations, it can be linked to mechanisms of anthropomorphism. These features are also called social cues (Mayer, 2014b; Schneider, Nebel, Pradel, & Rey, 2015a). Social cues activate a social response within learners and lead to an increase in active processing of the learning material (Mayer, 2014b). Since social cues are attributed as more humanlike (e.g., Woo, 2009), they also meet the definition of anthropomorphic features. For this, personalization can be seen as a form of anthropomorphic features. In addition, social cues enhance learners' interest and motivation while not increasing perceived cognitive load (e.g., Moreno & Mayer, 2004; Schneider, Nebel, Pradel, & Rey, 2015b). In an eye-tracking study by Zander, Reichelt, Wetzel, Kämmerer, and Bertel (2015), personalized learning materials have been proven to attract more attention through an increase of visual appeal. Moreover, Allen, Magnenat-Thalmann, and Thalmann (2012) have shown that time on task is significantly reduced for materials with personalization, although this study was conducted in order to examine the users' behavior in dense crowds of virtual environments.

Main Research Questions and Hypotheses

Based on the studies discussed in the previous sections, this experiment examines influences of anthropomorphic features within decorative pictures on learning outcomes of accompanied texts. More specifically, human faces were implemented in text-accompanied decorative pictures in order to enhance humanization of the whole learning material. This implementation was a common procedure of previous studies which enhanced learning performance via positive emotions (Um et al., 2012), a reduction of perceived difficulty (Sherman & Haidt, 2011), or attention-capturing (Park et al., 2015). However, the impact of this effect is not clear as learning pictures are exchanged by decorative pictures, which might be seductive and learning-hindering. For this, these anthropomorphized decorative pictures might become the center of attention and also distract learners from important information. In order to acknowledge both effect directions, two contrasting hypotheses are formulated:

> *Hypothesis 1a:* Learners who are shown decorative pictures including humanlike shapes will achieve *higher* learning scores than learners who are shown decorative pictures without additional anthropomorphic features.

> *Hypothesis 1b:* Learners who are shown decorative pictures including human-like shapes will achieve *lower* learning scores than learners who are shown decorative pictures without additional anthropomorphic features.

As well as human faces, personalization, as another feature of anthropomorphism, has been shown to enhance learning outcomes (Ginns et al., 2013) and attract learners' attention (Zander et al., 2015), when implemented within learning-relevant parts of the materials. In this experiment, picture labels of decorative pictures will be manipulated by their amount of personalization (nonpersonalized vs. personalized). Personalized labels may guide attention toward nonrelevant details. However, students might also become more familiar with a social situation (Schneider et al., 2015a) by this manipulation. Again, two contrasting hypotheses might be possible for this inclusion:

> *Hypothesis 2a:* Learners who are shown decorative pictures including personalized labels will achieve higher learning scores than learners who are shown decorative pictures with nonpersonalized labels.

> *Hypothesis 2b:* Learners who are shown decorative pictures including personalized labels will achieve lower learning scores than learners who are shown decorative pictures with nonpersonalized labels.

In a second experiment, fully anthropomorphized pictures (with humanlike shapes and personalized labels) were tested against nonanthropomorphized pictures (without humanlike shapes and personalized labels) and against a control group (no decorative pictures) in order to evaluate if anthropomorphisms increase, reduce, or even wipe out the seductive detail effect as supposed in the concept of conducive decorative pictures by Schneider et al. (2016).

> *Hypothesis 3a:* Learners who are shown anthropomorphized decorative pictures will achieve higher learning scores than

learners who are shown nonanthropomorphized decorative pictures and higher learning scores than learners without decorative pictures.

Hypothesis 3b: Learners who are shown anthropomorphized decorative pictures will achieve lower learning scores than learners who are shown nonanthropomorphized decorative pictures and lower learning scores than learners without decorative pictures.

In addition, as cognitive (e.g., mental effort; Mayer & Estrella, 2014), motivational (Plass et al., 2014) or affective variables (Um et al., 2012) have been shown to be influenced by anthropomorphic features, these aspects were additionally examined in both experiments.

Experiment 1

Method

Participants and design. The participants were 81 university students (68% female) from Chemnitz University of Technology, who received a 1-hr credit as a trial subject for their studies or 6 euro. The mean age was 25.20 years (SD = 4.10). Students are registered for subjects like psychology (29.6%), media and communication studies (27.2%), humanities (22.2%), economics and nature science (17.3%), and other studies (3.7%). Mean prior knowledge (further described in learning tasks) was 1.41 (SD = 1.29) out of 9 points.

This experiment aims at varying the amount of anthropomorphisms within decorative pictures via human faces and personalized labels. For this, each student was randomly assigned (block randomization) to one of the four experimental groups with decorative pictures of a two factorial between-subjects design. Accordingly, 21 students served in the human faces and personalized labeling group, 20 students in the human faces and nonpersonalized labeling group, 20 students in the without human faces and personalized labeling group, and 20 students in the without human faces and nonpersonalized labeling group. Since anthropomorphism is a new research field within multimedia learning and has not been pretested for decorative pictures, a possible operationalization via human faces in decorative pictures had to be pretested.

Materials and measures

Prestudy. In order to be able to implement decorative pictures with and without human faces, a prestudy has been conducted. Modifications of this form of anthropomorphism were accomplished through the addition of facelike structures like simplistic drawings of eyes and a mouth. Studies within developmental psychology have shown that even newborns recognize these features as human faces (e.g., Mondloch et al., 1999). However, main characteristics, such as a symmetry along the vertical axis or high contrast areas in the upper part of the drawing, must be complied with (Johnson & Morton, 1991). As the learning material of the main experiment (described in the Procedure section for Experiment 2) is about artificial intelligence (AI), 11 pictures of robots used in daily life (e.g., hospital robots, airport robots) were changed into an anthropomorphic and nonanthropomorphic version (Figure 1) by the inclusion or omission of the mentioned features and the exclusion of other human-like features (e.g., arms). Particular attention was paid to realistic counterparts of nonanthropomorphic, real-world, decorative pictures. As service robots, which are shown in the pictures, normally try to help people and imitate friendly characters, faces were depicted as smiling. Moreover, as studies of anthropomorphism within the field of emotional design show, anthropomorphic features are automatically connected with emotions (Park et al., 2015) and a positive valence rather enlarges the effect of anthropomorphism (Sherman & Haidt, 2011).

All 22 decorative picture were pretested on their perceived humanization. For this, 40 students (mean age: 24.00, SD = 3.82, 57.5% female) had to watch all pictures on web pages (one picture per page), and rate them on a 4-point scale according to their perceived humanization: "Please rate how human-like the shown robot is to you?" ranging from 0 (*not at all human-like*) to 3 (*very human-like*). Interrater-reliability can be seen as good, (ICC (2, k) = .844, $F(10, 790) = 6.39$, $p < .001$). All participants of this study were not allowed to take part in the main experiment. Results show that the group of pictures with anthropomorphic features (M = 1.49, SD = 1.10) was perceived as more humanlike than their contrast group (M = 0.50, SD = 1.10), $t(39) = 8.04$, $p < .001$, d = 1.14. In addition, pairwise t tests also revealed significant results for each picture comparison, $ps < .001$. These results clearly confirm that students are able to differentiate between anthropomorphic and nonanthropomorphic features within decorative pictures, so that this operationalization can be used within the main experiment.

Learning environment. The learning materials consisted of different texts (1,985 words) dividable in four main sections and 11 subsections. The texts described facts about AI (see Appendix A) based on scientific literature (Beckstein, 2014; Ertel, 2009). Each subsection was equipped with one decorative picture. Both texts and pictures are used to create four learning web pages (an overview is displayed in Figure 2). Participants had to navigate through all subpages to read all texts. The four titles of the main sections are displayed as buttons on the main page: "1. Introduction to AI," "2. AI and Perception," "3. AI and Awareness," and "4. AI as Search Method." In addition, a fifth button was displayed beneath the main section buttons, which was labeled with "I have read all texts." This button led participants to the subsequent questionnaires and learning tests and was not enabled unless all main sections had been read. When participants clicked on one of the main section buttons, they were shown subpages, which are all linked by a "Continue" button. Participants were instructed to navigate freely through all main sections. On the last subpage, this button leads back to the main page, where a green checkmark next to the corresponding main button indicated that this subsection has been read.

The decorative pictures displayed different kinds of service robots, which can be seen at public or private places (e.g., robotic vacuums, service robots in retirement homes). Since the learning text is about areas of AI (e.g., topics like "artificial neural nets," "the Turing test," or "the general problem solver"), these pictures do not convey learning-relevant information. These decorative pictures on each subpage differed among all experimental conditions according to the level of humanization (with or without human faces), or the degree of personalization within the picture labels (nonpersonalized vs. personalized labeling; for a comparison see Figure 1). The manipulation of human faces was taken

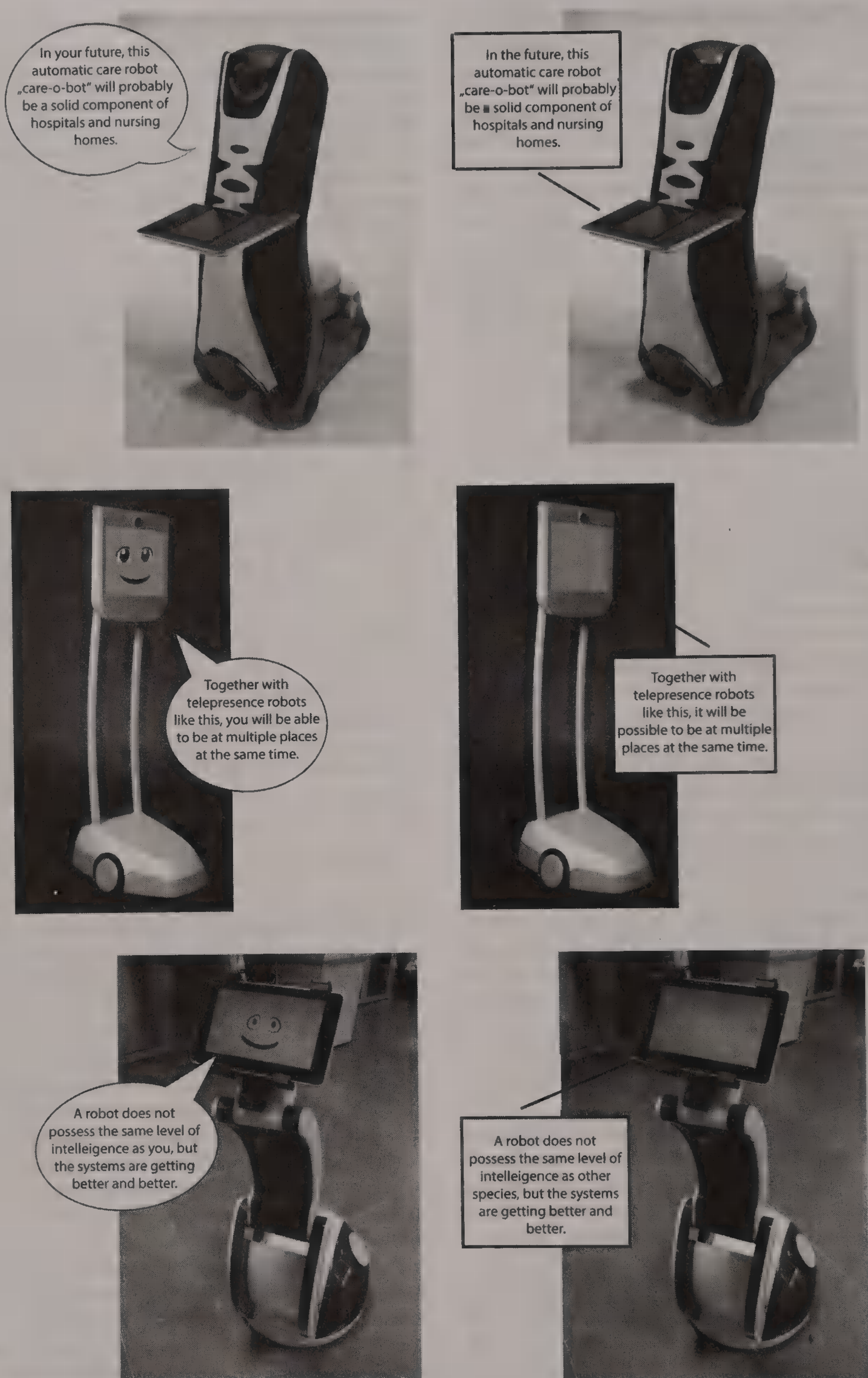

Figure 1. Comparison of three exemplary study pictures. Pictures with anthropomorphic and personalized features are in the left-hand column and pictures without any additional features are shown in the right-hand column. Labels are translated into English. See the online article for the color version of this figure.

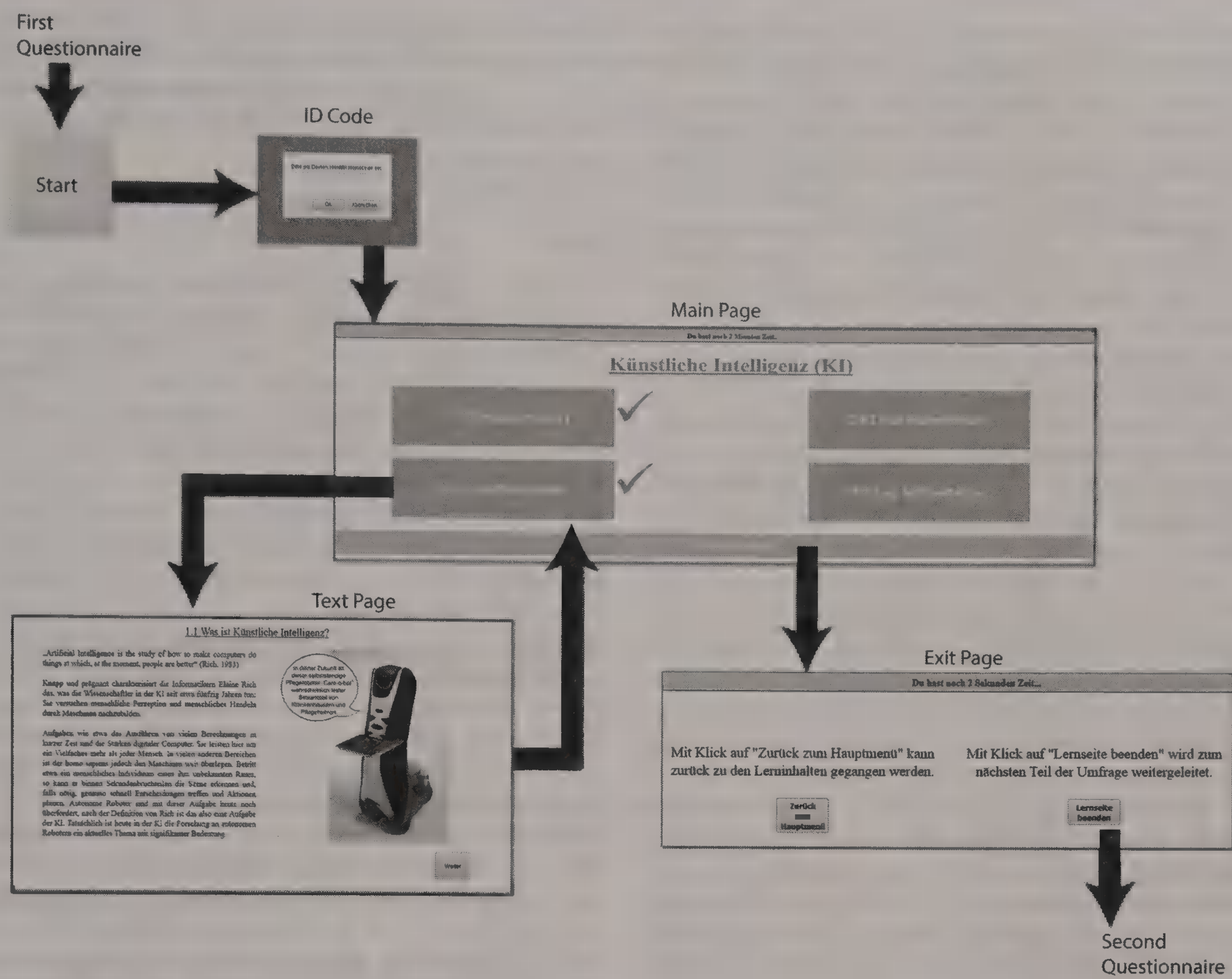

Figure 2. Overview of the learning environment with a selected text page. Arrows indicate possible click paths. See the online article for the color version of this figure.

from the prestudy. Personalization was implemented by a replacement of articles with personalized pronouns (e.g., "the" into "you," "your") within the speech bubbles (one article per label) instead of rectangular description boxes. This reflects a common operationalization of personalization (Ginns et al., 2013).

Independent from the experimental condition, a time bar was displayed at the top of all pages, which was fixed at 25 min. This amount of time was calculated through the mean of reading times by five nonexpert pretest readers, who have not read the text before. If participants of the main experiment were not able to read all texts in time, they were directed to the third part of the experiment. However, none of the readers were directed to the third part because of the time limit (M = 19.78 min, SD = 2.53). In addition, no significant main or interaction effects of the inclusion of human faces or types of labels concerning time ($p > .05$) can be observed.

Learning tests. Prior knowledge was measured with three open-answer questions: (a) "Define the term 'Artificial Intelligence'!"; (b) "What is a robot?"; (c) "What does the Turing-test mean?" These questions were based on the domain-specific knowledge which will be imparted by the following learning text. For each of the questions a preset answer catalog was prepared based on literature. For this, 3 points per questions could be reached—a sum of 9 points for all questions. Based on the answer catalog, answers were evaluated by two independent raters (ICC $(2, k) = .913$, $F(88, 88) = 11.55$, $p < .001$), who were not familiar with the experiment. The reliability score of all questions is acceptable ($\alpha = .70$) according to Cohen (1988).

In order to measure students' learning performance, retention and transfer scores, which are often used in multimedia learning studies (for an overview, see Mayer, 2014a), were created. According to Mayer (2014a), retention is defined as remembering. Remembering refers to being able to recognize or reproduce the learning content. Retention scores were determined by 10 questions on facts which can be found directly in the texts of the experiment (e.g., "How does the text define artificial intelligence?"). Since the first subpage was an introduction page, only 10 questions (one for each of the other subpages) were created. Each retention question was displayed with four possible answers. The number of correct answers differed among all tasks, however, at least one answer was correct. Each correct crossing or crossing-out was rewarded with 1 point. As a result, a maximum of four points was possible for each retention tasks and a maximum of 40 points was possible for all retention tasks.

The same procedure was used for all transfer tasks, whereby transfer problems are defined as understanding (Mayer, 2014a). Learners need a coherent mental model representation from the material in order to solve novel problems, which are not explicitly

presented in the learning material. For example, after reading the question: "Why is the Turing test important for spam blocking software?" students had to combine their knowledge on spam as a computerized software trying to imitate human e-mail correspondence and facts on the Turing test, which judges between human or machine behavior. Again, a maximum of 4 points can be reached for each transfer tasks and a maximum of 40 points for all transfer tasks. Regarding reliability, retention (α = .60) and transfer (α = .62) scores indicate a relatively stable internal consistency considering the multidimensionality of these constructs.

Additional questionnaires. One questionnaire was used to measure the emotional dimensions of students' valence (interrater reliability: α = .94) on a 7-point scale. The two items of this scale were derived from the Positive Affect, Negative Affect and VAlence Short Scales (PANAVA-KS) Questionnaire from Schallberger (2005). Students had to rate how they feel at the moment on the scales ranging from "unhappy" to "happy" and from "discontent" to "content." In order to assess students' mental effort, one item (i.e., "How much effort did you invest in understanding the learning material") from Paas (1992) was used. This item was accompanied with a 7-point scale ranging from *very low* to *very high*. Since decorative pictures are prone to distract learners from the learning materials, the task-irrelevant thinking scale (α = .88), derived from Sarason (1984), was included. The nine items of this scale (e.g., "During learning, irrelevant bits of information pop into my head") were displayed together with a 7-point scale ranging from *I totally disagree* to *I totally agree*. In order to measure intrinsic extrinsic motivation, the two scales of intrinsic motivation (α = .95) and external regulation (α = .86) from the Situational Motivation Questionnaire from Guay, Vallerand, and Blanchard (2000) were used. In this questionnaire students had to rate their motivation with the help of the question "Why are you currently engaged in this activity?" with the help of items like "Because I am supposed to do it."

Moreover, two manipulation check items were included. The first item was included to check if the anthropomorphic features are perceived as more human: "For me, the learning material was very human-like." The second item was implemented to analyze the personalization of the learning materials: "I feel personally touched by the learning material." Both items were displayed together with a 7-point scale ranging from *I totally disagree* to *I totally agree*. A demographic questionnaire was used to collect different demographic data, like age, sex or course of study.

Procedure. The study was conducted in a computer lab with 10 work stations. Students were randomly assigned to one of the four experimental groups by drawing lots and controlling the number of participants. Each accomplishment of an experiment consisted of one to four students. A corresponding number of computers had been prepared by opening the first experimental web page before each experiment started. All participants were instructed to follow the instructions on their screens, fill each gap within the questionnaires and read all information carefully. All students completed the three parts of the experiment autonomously, while the three parts of the experiment are connected via links. In the first part, prior knowledge was measured. In a second part, students had to navigate through the learning web pages. These learning web pages differed according to the experimental group. Within the third part all dependent variables were measured in the following order: (a) emotional, motivational, and cognitive scales; (b) learning tasks; and (c) manipulation check and demographic data. After students reached the last page, they needed to fill out a participants' list at the experimenter's table to reward them with 6 euro or a course credit. Overall, the experiment lasted 35 to 40 min.

Results

In the analysis of data, multivariate analyses of covariance (MANCOVAs) and follow-up univariate analyses of covariance (ANCOVAs) with human face and personalization as between-subjects factors were conducted in order to assess differences between groups. Predefined test assumptions are only reported if significant violations occurred. All analyses are corrected for prior knowledge and time on task as covariates, whereby only significant influences of the covariates were reported. There were no significant differences or interaction effects between the four treatment groups in terms of age, gender, reward type, subject, prior knowledge, or time on task (*ps* = [.073, .877]). Descriptive results of all dependent variables and covariates are displayed in Table 1.

Manipulation check. A manipulation check was conducted prior to the main analysis in order to ensure that the experimental manipulation among all independent variables succeeded. Thus, a MANCOVA was conducted with perceived humanlikeness and perceived personalization as dependent measures. Significant main effects were found for human faces, (Wilk's) $\Lambda = 0.37$, $F(2, 74) = 64.51$, $p < .001$, $\eta_p^2 = .64$, and for personalization, $\Lambda = 0.46$, $F(2, 74) = 43.95$, $p < .001$, $\eta_p^2 = .54$, but not for the interaction, $\Lambda = 0.97$, $F(2, 74) = 1.26$, $p = .291$, $\eta_p^2 = .03$.

An ANCOVA for perceived humanlikeness showed that pictures with human faces resulted in higher scores than pictures without these features, $F(1, 75) = 130.18$, $p < .001$, $\eta_p^2 = .63$, whereas no significant differences were found for personalization, $p = .850$, $\eta_p^2 = .01$. Regarding perceived personalization, a personalized seductive detail induced higher scores than pictures without any personalization, $F(1, 75) = 86.70$, $p < .001$, $\eta_p^2 = .53$. In contrast, no significant differences occurred between the manipulations of anthropomorphism, $p = .321$, $\eta_p^2 < .001$. Results show that both manipulations are recognized independently, so that further main effects and interaction can be evaluated properly.

Learning outcomes. In order to check possible influences of both factors on learning, a MANCOVA was conducted with retention and transfer scores as dependent measures. Significant main effects were found for human faces, $\Lambda = 0.75$, $F(2, 74) = 12.52$, $p < .001$, $\eta_p^2 = .25$, for personalization, $\Lambda = 0.84$, $F(2, 74) = 7.04$, $p = .002$, $\eta_p^2 = .16$, and for prior knowledge, $\Lambda = 0.98$, $F(2, 74) = 0.91$, $p = .036$, $\eta_p^2 = .09$, but not for the interaction, $\Lambda = 0.98$, $F(2, 74) = 0.89$, $p = .367$, $\eta_p^2 = .02$.

An ANCOVAs for retention revealed that pictures with human faces resulted in higher scores than pictures without these features, $F(1, 75) = 13.71$, $p < .001$, $\eta_p^2 = .16$. In contrast, no significant differences were found for personalization, $p > .05$, $\eta_p^2 = .001$. Results show that only faces in decorative pictures foster retention performance. Regarding transfer, a personalized seductive detail induced higher scores than pictures without any personalization, $F(1, 75) = 13.82$, $p < .001$, $\eta_p^2 = .15$ Additionally, significant differences occurred in favor of included human faces, $F(1, 75) = 10.62$, $p = .002$, $\eta_p^2 = .12$. No significant interaction was found, $p > .05$, $\eta_p^2 = .02$. Both faces in decorative pictures and person-

Table 1
Mean Scores of All Dependent Variables and Prior Knowledge Together With Their Corresponding Standard Deviations for the Four Experimental Picture Groups of Experiment 1

	Experimental groups							
	With human faces				Without human faces			
	Personalized labels (N = 21)		Nonpersonalized labels (N = 20)		Personalized labels (N = 20)		Nonpersonalized labels (N = 20)	
Type of scale	M	SD	M	SD	M	SD	M	SD
Perceived anthropomorphism	5.38	.86	4.70	.92	2.45	1.23	2.55	1.00
Perceived personalization	5.00	1.00	2.90	.97	5.10	.97	2.95	1.00
Prior knowledge	1.62	1.47	.85	.81	1.70	1.22	1.45	1.47
Time on task (min)	19.55	2.43	20.14	2.85	19.85	2.06	19.59	2.85
Retention	29.52	2.52	28.75	3.45	26.30	3.18	27.00	3.66
Transfer	29.86	2.57	28.65	2.83	28.50	2.80	25.70	2.85
Task-irrelevant thinking	3.60	1.17	3.72	1.44	3.08	1.31	2.66	1.14
Mental effort	5.05	.92	4.40	1.05	4.05	1.10	3.20	1.24
Intrinsic motivation	4.71	.76	4.69	1.02	4.29	.81	4.03	.98
External regulation	4.03	1.58	3.80	1.78	3.89	1.86	3.33	1.54
Valence	4.78	.70	4.98	.73	3.68	.88	3.65	.90

Note. The scores of perceived anthropomorphism, perceived personalization, mental effort, task-irrelevant thinking, intrinsic motivation, external regulation and valence ranged from 1 to 7. Prior knowledge ranged from 0 to 9, whereas retention and transfer scores ranged from 0 to 40.

alized labels were found to foster transfer performance. Retentions scores did not significantly change due to prior knowledge (p = .207), however, transfer scores were affected by this covariate, $F(1, 75) = 5.93$, $p = .017$, $\eta_p^2 = .06$. Taken together, these results fully confirm hypothesis H_{1a} and partially confirm hypothesis H_{2a}, while H_{1b} and H_{2b} can be rejected.

Cognitive variables. In order to check influences of both experimental factors on cognitive processes, a MANCOVA was conducted with task-irrelevant thinking and mental effort scores as dependent measures. Significant main effects were found for human faces, $\Lambda = 0.71$, $F(2, 74) = 15.00$, $p < .001$, $\eta_p^2 = .29$, for personalization, $\Lambda = 0.87$, $F(2, 74) = 5.71$, $p = .005$, $\eta_p^2 = .13$, and for time on task, $\Lambda = 0.90$, $F(2, 74) = 4.13$, $p = .030$, $\eta_p^2 = .10$, but not for the interaction, $\Lambda = 0.98$, $F(2, 74) = 0.60$, $p > .05$, $\eta_p^2 = .02$.

Concerning task-irrelevant thinking, pictures with human faces resulted in higher scores than pictures without these features, $F(1, 75) = 8.33$, $p = .005$, $\eta_p^2 = .10$, whereas no significant differences can be found for personalization, $p = .701$, $\eta_p^2 = .002$. Regarding mental effort, personalized labels in decorative picture induced higher scores than pictures without personalization, $F(1, 75) = 11.21$, $p = .001$, $\eta_p^2 = .13$. In addition, significant differences occurred between the manipulation of human faces, $F(1, 75) = 20.27$, $p < .001$, $\eta_p^2 = .21$, whereby human faces increased mental effort. No significant interactions can be found, $ps > .382$, $\eta_p^2 < .01$. Mental effort results significantly changed according to the covariate ($p = .007$, $\eta_p^2 = .09$). In sum, both anthropomorphism features were found to increase the invested mental effort. In contrast, only faces in decorative pictures increase the amount of irrelevant thoughts during learning.

Motivational and affective processes. In order to check the influences of both experimental factors on affective and motivational processes, a MANCOVA was conducted with intrinsic motivation, extrinsic motivation, and valence scores as dependent measures. Significant effects were found for human faces, $\Lambda = 0.53$, $F(2, 74) = 22.04$, $p < .001$, $\eta_p^2 = .48$, and for time on task, $\Lambda = 0.85$, $F(2, 74) = 4.23$, $p = .008$, $\eta_p^2 = .15$, but neither for personalization, $\Lambda = 0.98$, $F(2, 74) = 0.42$, $p = .739$, $\eta_p^2 = .02$, nor the interaction, $\Lambda = 0.97$, $F(2, 74) = 0.65$, $p = .584$, $\eta_p^2 = .03$.

An ANCOVA for intrinsic motivation revealed that pictures with human faces caused higher scores than pictures without these features, $F(1, 75) = 9.53$, $p = .003$, $\eta_p^2 = .11$. Regarding extrinsic motivation, faces did not generate significant differences, $p > .05$, $\eta_p^2 < .01$. Results of valence show that pictures with human faces caused higher scores than pictures without these features, $F(1, 75) = 48.42$, $p < .001$, $\eta_p^2 = .39$. Intrinsic motivation scores significantly changed according to the covariate time on task ($p = .003$, $\eta_p^2 = .11$). Faces in decorative pictures were found to increase the intrinsic motivation and positive valence of students.

Conclusion

This experiment was designed in order to evaluate if learning-enhancing mechanisms of anthropomorphism will meliorate or deteriorate the impact of decorative pictures. Results show that anthropomorphism (human faces and personalized labels) was able to enhance both learning scores retention and transfer. Personalized pictures were not able to foster retention but transfer scores. This effect pattern of fostering transfer rather than retention features is consistent with previous literature concerning the personalization effect (Ginns et al., 2013; Schworm & Stiller, 2012). Overall, these results show that implementing learning-enhancing features into decorative pictures are able to foster memory and knowledge transfer processes.

Discussion and Limitations

A closer look into possible processes behind the inclusion of anthropomorphism shows that human faces were able to increase a positive valence of the students. Since all of the included features show smiling faces, this might be a common result. However, students also perceived a higher intrinsic motivation which might

be traceable to a higher situational interest of the students (Bye, Pushkar, & Conway, 2007) and to the higher assessed positive valence (Pekrun, Goetz, Titz, & Perry, 2002). The increase of positive valence might also enable students to selectively attend to goal-relevant information and elaboration strategies (Pekrun et al., 2002), which, in turn, has increased participants' behavior to actively process the learning material as shown by increasing mental effort scores. However, students also reported higher scores of task-irrelevant thinking. These increased scores might be explained by a strong cognitive linkage of anthropomorphism and prior experiences, such as thoughts of the last smiling person. According to the equal emotional states of both the learning material and the thoughts about past events, anthropomorphism might have functioned as a process of empathy (Airenti, 2015). This link might also be the reason that students were able to better retrieve information because more cognitive access points to the important information can be used.

Within the manipulation of personalization, direct speech is assessed in combination with higher scores of mental effort in contrast to formal speech. This finding is consistent with previous research (Kurt, 2011). In contrast to other studies, personalized labels were not able to enhance motivation or affect and none of the implemented attention-attracting features were able to influence the perception of extrinsic motivation. In sum, it can be argued that anthropomorphism is shown to sustainably enhance motivation since persistence-facilitating intrinsic motivation, in particular, is enhanced rather than a short-termed extrinsic motivation (Vansteenkiste et al., 2004).

Experiment 2

A second experiment was conducted to investigate if anthropomorphized decorative pictures, in general, can be seen as rather conducive (Schneider et al., 2016) or seductive (Harp & Mayer, 1998). For this, a control group without decorative pictures is needed. Since motivation and emotion scores are not baseline adjusted, these scores need to be included before and after the learning material. A closer look at possible differences in cognitive facets (intrinsic, extraneous, and germane cognitive load; Kalyuga, 2011) would additionally help to evaluate the learning process, since decorative pictures are supposed to increase extraneous processing, while anthropomorphism rather decreases perceived difficulties (intrinsic cognitive load) and facilitates germane cognitive load. For this, Experiment 2 was conducted in order to substantiate the results of Experiment 1.

Method

Participants and design. The participants were 102 school students (48.5% female) from a secondary school in Thuringia. The mean age was 14.39 years (SD = 1.17). Students were recruited from Class Levels 8 (56.3%) and 9. Mean prior knowledge was 0.40 (SD = 0.77) out of 9 points. The experiment was conducted in computer science classes.

This experiment aims at varying the amount of anthropomorphisms within decorative pictures. For this, 34 students were randomly assigned (block randomization) to each of the three experimental groups (anthropomorphized, nonanthropomorphized, without decorative pictures). One material included anthropomorphized decorative pictures. This group of pictures included human faces and personalized labels (as described in Experiment 1). A second material included pictures without anthropomorphized features. This group received the same pictures as those without faces and personalized labels used in Experiment 1. In addition, a learning material without decorative pictures was developed in order to evaluate if the other groups can be seen as rather seductive or conducive.

Materials and measures.

Learning environment. The same learning web pages from Experiment 1 were used in this experiment except for the used decorative pictures. These pictures differed according to the experimental groups. Again, none of the readers was directed to the questionnaires after the learning materials because of the time limit (M = 17.79 min, SD = 2.93). In addition, no significant differences concerning time ($p > .05$) occurred between all group comparisons.

Questionnaires and tests. The same prior knowledge (α = .71), retention (α = .69), and transfer (α = .76) questionnaires as used in Experiment 1 were taken for this experiment. In addition, the intrinsic and extrinsic motivation scales as well as the valence scale from Experiment 1 were included. As proposed after Experiment 1, a closer look at cognitive processes would be helpful to explain how anthropomorphism results in a higher learning performance. For this, the cognitive load questionnaire by Leppink, Paas, van Gog, van Der Vleuten, and van Merriënboer (2014) was included. This questionnaire contains three scales along an 11-point scale measuring perceived intrinsic (ICL; α = .89), extraneous (ECL; α = .79) and germane (GCL; α = .95) cognitive load scores. Example items are "The topics covered in the lecture were very complex" (ICL), "The instructions and explanations during the lecture were very unclear" (ECL), and "The lecture really enhanced my understanding of the topics covered" (GCL). These items were adapted to the text-based environment (see Appendix B). A manipulation check and a demographic questionnaire were included as described in Experiment 1, except that subject of study was exchanged by class level.

Procedure. The study was conducted in a computer lab of a school with 25 work stations. Students were randomly assigned to one of the three experimental groups by drawing lots, and controlling for the number of participants. Each accomplishment of an experiment consisted of seven to 11 students. A corresponding number of computers had been prepared by opening the first experimental web page before each experiment started. All participants were instructed to follow the instructions on their screens, fill each gap within the questionnaires and read all information carefully. All students completed the three parts of the experiment individually, while the three parts of the experiment are connected via links. In the first part, prior knowledge was measured. In addition, all motivational and emotional questionnaires were included in order to secure control measurements before the learning environment. In a second part, students had to navigate through the learning web pages. These learning web pages differed according to the experimental group. Within the third part all dependent variables were measured in the following order: (a) emotional, motivational, and cognitive load scales; (b) learning tasks; and (c) manipulation check and demographic data. After students reached the last page, they were instructed to sit quietly at their work stations. At the end of the lesson, students needed to fill out a

participants' list at the teacher's table to reward each of them with a chocolate bar. Overall, the experiment lasted 40 to 50 min.

Results

In the analysis of data, MANCOVAs and follow-up univariate ANCOVAs with human face and personalization as between-subjects factors were conducted in order to assess differences between groups. Predefined test assumptions are only reported if significant violations occurred. All analyses are corrected for prior knowledge and time on task as covariates, whereby only significant influences of the covariates are reported. There were no significant differences or interaction effects between the four treatment groups in terms of age, gender, class level, prior knowledge, or time on task (ps = [.431, .962]). In addition, ANCOVAs for the a priori measure of valence, $F(2, 99) = 2.39$, $p = .088$, $\eta_p^2 = .04$, intrinsic motivation, $F(2, 99) = 1.34$, $p = .268$, $\eta_p^2 = .03$, and external regulation, $F(2, 99) = 0.24$, $p = .787$, $\eta_p^2 = .01$, did not reveal significant differences. Descriptive results of all dependent variables and covariates are displayed in Table 2.

Manipulation check. Again, a manipulation check was conducted prior to the main analysis in order to check if the manipulation between both picture groups succeeded. Thus, a MANCOVA was conducted using perceived humanlikeness and perceived personalization as dependent measures. A significant main effect was found for picture groups, (Wilk's) $\Lambda = 0.75$, $F(2, 62) = 10.21$, $p < .001$, $\eta_p^2 = .25$. An ANCOVA for perceived human-likeness revealed that anthropomorphized pictures caused higher scores than pictures without these features, $F(1, 63) = 10.68$, $p = .002$, $\eta_p^2 = .15$. Regarding perceived personalization, anthropomorphized pictures induced higher scores than pictures without any personalization, $F(1, 63) = 19.25$, $p < .001$, $\eta_p^2 = .23$. With these results, the manipulation can be seen as accepted.

Learning outcomes. In order to check possible differences between all three groups among learning outcomes, a MANCOVA was conducted with retention and transfer scores as dependent measures. Significant effects were found between groups, $\Lambda = 0.66$, $F(2, 96) = 10.90$, $p < .001$, $\eta_p^2 = .19$. An ANCOVA for retention revealed a significant effect for the group differences, $F(2, 97) = 18.14$, $p < .001$, $\eta_p^2 = .27$. Regarding transfer, significant differences were also shown, $F(2, 97) = 7.64$, $p = .001$, $\eta_p^2 = .14$.

Subsequently, Bonferroni-Holm-corrected pairwise comparisons for all learning scores were conducted. Regarding retention, learners with anthropomorphized pictures scored significantly higher than learners without pictures (mean difference = 2.57, $p = .001$, $\eta_p^2 = .15$), and higher than learners with nonanthropomorphized pictures (mean difference = 4.68, $p < .001$, $\eta_p^2 = .36$). In addition, the control group reached significantly higher scores than the group with nonanthropomorphized pictures (mean difference = 2.11, $p = .008$, $\eta_p^2 = .10$). Concerning transfer, learners with anthropomorphized pictures significantly outperformed learners without pictures (mean difference = 2.50, $p = .002$, $\eta_p^2 = .13$) and also learners with nonanthropomorphized pictures (mean difference = 2.96, $p < .000$, $\eta_p^2 = .16$). In contrast to the retention results, there was no significant difference between the control group and the group with nonanthropomorphized pictures (mean difference = 0.47, $p = .58$). Taken together, results mainly confirm hypothesis H_{3a}.

Cognitive processes. In order to check the influences of both experimental factors on cognitive processes, a MANCOVA was conducted with the ICL, ECL, and GCL scores as dependent measures. A significant main effect was found for group, $\Lambda = 0.63$, $F(6, 190) = 8.37$, $p < .001$, $\eta_p^2 = .21$. ANCOVAS for ICL, $F(2, 97) = 3.94$, $p = .023$, $\eta_p^2 = .08$; ECL, $F(2, 97) = 6.90$, $p = .002$,

Table 2

Mean Scores of All Dependent Variables and Covariates Together With Their Corresponding Standard Deviations for the Three Experimental Groups of Experiment 2

	Experimental groups					
	Anthropomorphized pictures (N = 34)		Nonanthropomorphized pictures (N = 34)		No pictures (N = 34)	
Type of scale	M	SD	M	SD	M	SD
Perceived anthropomorphism	3.79	1.39	2.73	1.31		
Perceived personalization	3.88	1.12	2.52	1.30		
Prior knowledge	.44	.89	.26	.62	.50	.79
Time on task (min)	18.22	2.80	17.48	3.24	17.66	2.75
Retention	22.32	3.49	17.65	3.14	19.71	2.80
Transfer	24.59	3.39	21.59	3.34	22.03	3.21
Intrinsic cognitive load	4.73	1.59	5.32	1.85	5.98	1.83
Extraneous cognitive load	5.00	1.43	5.09	1.71	3.86	1.53
Germane cognitive load	6.21	1.73	4.41	1.78	6.09	2.08
Intrinsic motivation (before)	3.10	1.06	3.36	1.27	3.54	1.07
Intrinsic motivation (after)	3.86	1.13	3.24	1.15	3.39	1.27
External regulation (before)	4.01	1.42	4.05	2.00	3.80	1.32
External regulation (after)	4.07	1.31	4.04	1.68	4.99	1.26
Valence (before)	4.07	1.07	4.59	1.08	4.44	.76
Valence (after)	4.75	1.12	3.82	1.30	3.60	1.19

Note. "Before" means a measurement before the learning phase and "after" means a measurement after the learning phase. The scores of perceived anthropomorphism, perceived personalization, intrinsic motivation (before and after), external regulation (before and after), and valence (before and after) ranged from 1 to 7. Intrinsic cognitive load, extraneous cognitive load and germane cognitive load ranged from 1 to 11. Prior knowledge ranged from 0 to 9, whereas retention and transfer scores ranged from 0 to 40.

$\eta_p^2 = .13$; and GCL, $F(2, 97) = 9.48$, $p < .001$, $\eta_p^2 = .16$, showed significant effects.

Subsequently, Bonferroni-Holm-corrected pairwise comparisons for all perceived cognitive load scores were conducted. Only learners with anthropomorphized pictures assessed their learning material as significantly lower in ICL than learners without pictures (mean difference = 1.26, $p = .004$, $\eta_p^2 = .11$). No other group comparisons differ significantly (mean difference = [0.52; 0.67], $p >$ [.120; .227]). Learners with anthropomorphized pictures assessed their learning material as significantly higher in ECL than learners without pictures (mean difference = 1.19, $p = .002$, $\eta_p^2 = .12$), but not significantly higher than learners with nonanthropomorphized pictures (mean difference = 0.10, $p = .951$). Students in the control group rated their ECL also lower than the group with nonanthropomorphized pictures (mean difference = 1.22, $p = .002$, $\eta_p^2 = .14$). In addition, learners with anthropomorphized pictures reported their GCL significantly higher than learners with nonanthropomorphized pictures (mean difference = 1.80, $p < .001$, $\eta_p^2 = .18$) but not higher than learners in the control group (mean difference = 0.14, $p = .680$). Moreover, nonanthropomorphized pictures were assessed as significantly lower in GCL than the control group (mean difference = 1.61, $p = .001$, $\eta_p^2 = .16$).

Motivational and affective processes. Since the intrinsic motivation, extrinsic motivation, and valence scores were measured directly before and after the learning environment, difference scores (post hoc minus a priori measurement) for each scale (valence, intrinsic motivation, and external regulation) were calculated and used in a subsequent MANCOVA. Significant effects were found between groups, $\Lambda = 0.68$, $F(6, 190) = 6.88$, $p < .001$, $\eta_p^2 = .18$. Follow-up ANCOVAs for valence difference, $F(2, 97) = 11.52$, $p < .001$, $\eta_p^2 = .19$; intrinsic motivation difference, $F(2, 97) = 8.45$, $p < .001$, $\eta_p^2 = .15$; and external regulation difference, $F(2, 97) = 6.789$, $p = .002$, $\eta_p^2 = .12$, revealed a significant effect for the group differences.

Bonferroni-Holm-corrected pairwise comparisons for valence difference showed that learners with anthropomorphized pictures assessed their intrinsic motivation as significantly higher than learners without pictures (mean difference = 1.52, $p < .001$, $\eta_p^2 = .22$), and significantly higher than learners with nonanthropomorphized pictures (mean difference = 1.44, $p < .001$, $\eta_p^2 = .20$). In addition, there was no significant difference between the control group and the group with nonanthropomorphized pictures (mean difference = 0.07, $p = .836$).

Bonferroni-Holm-corrected pairwise comparisons for intrinsic motivation difference showed that learners with anthropomorphized pictures assessed their intrinsic motivation as significantly higher than learners without pictures (mean difference = 0.93, $p < .001$, $\eta_p^2 = .17$), and significantly higher than learners with nonanthropomorphized pictures (mean difference = 0.88, $p = .001$, $\eta_p^2 = .15$). In addition, there was no significant difference between the control group and the group with nonanthropomorphized pictures (mean difference = 0.05, $p = .256$).

Bonferroni-Holm-corrected pairwise comparisons external regulation difference showed that learners with anthropomorphized pictures assessed their external regulation as significantly lower than learners without pictures (mean difference = 1.13, $p = .003$, $\eta_p^2 = .12$), but not significantly lower than learners with nonanthropomorphized pictures (mean difference = 0.11, $p = .770$). In addition, the control group was assessed as significantly higher than the group with nonanthropomorphized pictures (mean difference = 1.24, $p = .001$, $\eta_p^2 = .14$).

General Discussion

Results of Experiment 2 demonstrated that the participants within the anthropomorphism condition outperformed the control group and the nonanthropomorphized pictures condition regarding knowledge retention. Additionally, as the control condition revealed larger retention scores than the nonanthropomorphized pictures group, the seductive detail effect was replicated in this case. Regarding transfer knowledge, the anthropomorphism condition outperformed all other experimental groups. In contrast to retention, no difference between the control group and the nonanthropomorphized pictures condition was found. This is especially interesting as the manipulation of personalization in Experiment 1 revealed differences between students' transfer scores (i.e., higher scores for personalized labels). Furthermore, to enrich the ongoing debate regarding seductive or conducive pictures (e.g., Schneider et al., 2016), the results demonstrated a successful separation of different types of pictures regarding their impacts on learning.

To extend the findings of Experiment 1 in relation to cognitive processes (a higher amount of task-irrelevant-thinking for human faces and higher scores of mental effort in both anthropomorphism variations), Experiment 2 included a measurement based on cognitive load. Consistent with the observations of more invested mental effort, the inclusion of decorative pictures increased ECL. However, the inclusion of anthropomorphized pictures reduced the ICL in contrast to the control group, whereas the nonanthropomorphized pictures condition was not significantly different from the control group. Additionally, participants within the anthropomorphized pictures group rated their GCL higher than participants within the nonanthropomorphized group. Finally, the control group demonstrated higher GCL scores than the nonanthropomorphized pictures group. These detailed results enable a deeper look into the mechanisms of picture procession within learning materials. Decorative pictures might trigger the seductive detail effect (higher ECL) in general (Lehman, Schraw, McCrudden, & Hartley, 2007). At the same time, anthropomorphism was shown to lower a perceived difficulty (lower ICL) and increase relevant learning processes (higher GCL). In contrast, the nonanthropomorphized pictures do not trigger these useful mechanisms. These results show that anthropomorphic features moderate the seductive detail effect.

As Experiment 1 only revealed improved affect and motivation with human faces, but not with the manipulation of personalization, Experiment 2 was needed to validate if a simultaneous manipulation could show the same patterns. Results revealed higher intrinsic motivation scores within the anthropomorphized pictures group compared to all other experimental groups. Additionally, the control group demonstrated higher scores of extrinsic motivation than the conditions with pictures. This underlines the results of Experiment 1. Partially contrasting with the valence results of Experiment 1, Experiment 2 revealed that only participants within the anthropomorphized pictures condition achieved significantly higher scores. Therefore, this experiment underpins that not only the inclusion of pictures but also how they are

designed (anthropomorphized or not) impacts affect and motivation.

Implications

The results of this study enlarge research on personalization and anthropomorphism in multimedia learning and enhance research on decorative pictures by providing an opportunity to dampen the seductive detail effect. The current investigation also demonstrates that it is possible to evoke positive affective and motivational states of learners by including decorative pictures. This is an alternative which could be easier to implement in multimedia learning environments than videos with emotional content (Plass et al., 2014) or a general aesthetical design (Heidig, Müller, & Reichelt, 2015). The current experiment outlines that an anthropomorphic design of decorative pictures has a positive impact on learning processes. This result supports the "focused more is more" approach, postulated by Mayer (2014a), which states that additional decorative elements that aim at increasing motivational states should be implemented in a modest way. Under several conditions, additional information can foster learning. The current investigation also underlines the role of affective and motivational processes in learning with multimedia, as shown in the ICALM, in addition to purely cognitive learning influences (Mayer, 2014a). In addition, anthropomorphism can be integrated in the research on possible emotional design elements (e.g., Park et al., 2015).

On the practical side, the results of this study suggest that designers of learning material should be encouraged to make use of decorative pictures in the awareness of the moderating role of anthropomorphism. In addition, decorative pictures with a high level of anthropomorphism should be implemented in learning materials in order to evoke positive affective states and strengthen (intrinsic) motivational processes.

Limitations

In the context of the present experiment, only short-term effects of decorative pictures in learning environments on learning with multimedia can be interpreted. Although strong effects were found, it does not signify that they are stable over a certain period of time. Since learning was self-paced and participants were able to choose freely how often they wanted to read texts, this may have compensated for any detrimental effects of decorative pictures: for example, a reduced attention allocation to the text (e.g., Sung & Mayer, 2012). Anthropomorphism was operationalized inter alia by adding friendly facelike structures to pictures of robots in order to reflect realistic situations. Because of the expression of positive emotion, effects of anthropomorphism and valence are not clear cut, however, this could enhance a more realistic picture of anthropomorphism. Multiple measurement of all cognitive, affective, or motivational variables at different points of time could have given more insight into possible mechanisms of the effects (e.g., Rop, van Wermeskerken, de Nooijer, Verkoeijen, & van Gog, 2016). Since the separation of cognitive load facets is still discussed controversially (Leppink et al., 2014), interpretation needs to be drawn in light of this discussion. This study incorporates one learning topic only in order to strengthen the comparability of both experiments, however, this might reduce a possible generalization to other topics. Eye-tracking data should be examined in further investigations since anthropomorphism was found to attract attention. In addition, decorative pictures differ in their amount of relevance to the learning goal compared with other studies (e.g., Harp & Mayer, 1998). This difference, however, was not measured, so that a clear connection toward seductive detail studies is limited.

Future Directions

As our research answers several research questions, other challenges emerge. As we have used the same learning material in order to make results more comparable, this connection should be applied to other learning materials in order to be able to generalize findings even more. For example, does the extent of anthropomorphism within virtual learning environments evoke similar effects? Another interesting area of further research can be derived from results regarding perceived irrelevant thinking and extraneous cognitive load. Although both scores increased in the anthropomorphized pictures groups and therefore should have dampened the learning outcomes, learning performance increased. Maybe the increased mental effort or a reduced perceived difficulty (ICL) counterbalanced negative effects. Possible interactions should be researched further. Finally, as we demonstrated the use of anthropomorphism to influence mood, the inclusion of such features as a method of mood management within multimedia learning material should be considered.

References

Airenti, G. (2015). The cognitive bases of anthropomorphism: From relatedness to empathy. *International Journal of Social Robotics, 7,* 117–127. http://dx.doi.org/10.1007/s12369-014-0263-x

Allen, B. F., Magnenat-Thalmann, N., & Thalmann, D. (2012). Politeness improves interactivity in dense crowds. *Computer Animation and Virtual Worlds, 23,* 569–578. http://dx.doi.org/10.1002/cav.1472

Beckstein, C. (2014). Suche [Search]. In G. Görtz, J. Schneeberger, & U. Schmidt (Eds.), *Handbuch der künstlichen intelligenz* (pp. 75–104). München, Germany: Oldenbourg.

Bye, D., Pushkar, D., & Conway, M. (2007). Motivation, interest, and positive affect in traditional and nontraditional undergraduate students. *Adult Education Quarterly, 57,* 141–158. http://dx.doi.org/10.1177/0741713606294235

Carney, R. N., & Levin, J. R. (2002). Pictorial pictures still improve students' learning from text. *Educational Psychology Review, 14,* 5–26. http://dx.doi.org/10.1023/A:1013176309260

Chen, X., & Latham, G. P. (2014). The effect of priming learning vs. performance goals on a complex task. *Organizational Behavior and Human Decision Processes, 125,* 88–97. http://dx.doi.org/10.1016/j.obhdp.2014.06.004

Chiaverina, C., Scott, C., & Steele, P. (1997). The connections project: Art, physics, and mathematics. *The Physics Teacher, 35,* 292–294. http://dx.doi.org/10.1119/1.2344688

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Ertel, W. (2009). *Grundkurs Künstliche Intelligenz; Eine praxisorientierte Einführung* [Introduction to artificial intelligence; a practical guide]. Wiesbaden, Germany: Vieweg+Teubner.

Farah, M. J., & Heberlein, A. S. (2007). Personhood and neuroscience: Naturalizing or nihilating? *The American Journal of Bioethics, 7,* 37–48. http://dx.doi.org/10.1080/15265160601064199

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social*

Psychology, 60, 509–517. http://dx.doi.org/10.1037/0022-3514.60.4.509

Ginns, P., Martin, A. J., & Marsh, H. W. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review, 25,* 445–472. http://dx.doi.org/10.1007/s10648-013-9228-0

Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior, 24,* 1494–1509. http://dx.doi.org/10.1016/j.chb.2007.05.007

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion, 24,* 175–213. http://dx.doi.org/10.1023/A:1005614228250

Haaranen, L., Ihantola, P., Sorva, J., & Vihavainen, A. (2015). In search of the emotional design effect in programming. In P. Kellenberger (Ed.), *Proceedings of the 37th International Conference on Software Engineering* (Vol. 2, pp. 428–434). Hoes Lane, NJ: IEEE.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90,* 414–434. http://dx.doi.org/10.1037/0022-0663.90.3.414

Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior, 44,* 81–95. http://dx.doi.org/10.1016/j.chb.2014.11.009

Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition.* Oxford, UK: Basil Blackwell.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23,* 1–19. http://dx.doi.org/10.1007/s10648-010-9150-7

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior, 28,* 241–250. http://dx.doi.org/10.1016/j.chb.2011.09.006

Kurt, A. A. (2011). Personalization principle in multimedia learning: Conversational versus formal style in written word. *Turkish Online Journal of Educational Technology, 10,* 185–192.

Lee, E. J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior, 26,* 665–672. http://dx.doi.org/10.1016/j.chb.2010.01.003

Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology, 32,* 569–587. http://dx.doi.org/10.1016/j.cedpsych.2006.07.002

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science, 41,* 811–831. http://dx.doi.org/10.1007/s11251-012-9256-z

Leppink, J., Paas, F., van Gog, T., van Der Vleuten, C. P., & van Merrienboer, J. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30,* 32–42. http://dx.doi.org/10.1016/j.learninstruc.2013.12.001

Levie, W. H., & Lentz, R. (1982). Effects of text pictures: A review of research. *Educational Communication and Technology: A Journal of Theory, Research, and Development, 30,* 195–232.

Magner, U. I., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering situational interest by illustrations pictures both fosters and hinders learning in computer-based learning environments. *Learning and Instruction, 29,* 141–152. http://dx.doi.org/10.1016/j.learninstruc.2012.07.002

Male, A. (2007). *Picture: A theoretical and contextual perspective.* Lausanne, Switzerland: Ava Book.

Mayer, R. E. (2001). *Multimedia learning.* New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139164603

Mayer, R. E. (Ed.). (2014a). *The Cambridge handbook of multimedia learning.* New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369

Mayer, R. E. (Ed.). (2014b). Principles of multimedia learning based on social cues: Personalization, voice, and image principles. *The Cambridge handbook of multimedia learning* (pp. 345–370). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.017

Mayer, R. E., & Estrella, G. (2014). Benefits of emotional design in multimedia instruction. *Learning and Instruction, 33,* 12–18. http://dx.doi.org/10.1016/j.learninstruc.2014.02.004

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science, 10,* 419–422. http://dx.doi.org/10.1111/1467-9280.00179

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning, 22,* 149–158.

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96,* 165–173. http://dx.doi.org/10.1037/0022-0663.96.1.165

Nowak, K. L., Hamilton, M. A., & Hammond, C. C. (2009). The effect of image features on judgments of homophily, credibility, and intention to use as avatars in future interactions. *Media Psychology, 12,* 50–76. http://dx.doi.org/10.1080/15213260802669433

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84,* 429–434. http://dx.doi.org/10.1037/0022-0663.84.4.429

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education, 86,* 30–42. http://dx.doi.org/10.1016/j.compedu.2015.02.016

Park, S., Kim, M., Lee, Y., Son, C., & Lee, M. (2005). The effects of visual pictures on learners' achievement and interest in PDA-(personal digital assistant) based learning. *Journal of Educational Computing Research, 33,* 173–187. http://dx.doi.org/10.2190/ELEY-NPQN-0WVM-FW0G

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37,* 91–105. http://dx.doi.org/10.1207/S15326985EP3702_4

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction, 29,* 128–140. http://dx.doi.org/10.1016/j.learninstruc.2013.02.006

Plass, J. L., & Kaplan, U. (2015). Emotional design in digital media for learning. In S. Tettegah, & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–162). New York, NY: Elsevier.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7,* 216–237. http://dx.doi.org/10.1016/j.edurev.2012.05.003

Rop, G., van Wermeskerken, M., de Nooijer, J. A., Verkoeijen, P. P., & van Gog, T. (2016). Task experience as a boundary condition for the negative effects of irrelevant information on learning. *Educational Psychology Review.* Advance online publication. http://dx.doi.org/10.1007/s10648-016-9388-9

Rubens, P. (Ed.), (2000). *Science and technical writing: A manual of style.* London, UK: Routledge.

Samuels, S. J. (1970). Effects of pictures on learning to read, comprehension and attitudes. *Review of Educational Research, 40,* 397–407. http://dx.doi.org/10.3102/00346543040003397

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46,* 929–938. http://dx.doi.org/10.1037/0022-3514.46.4.929

Schallberger, U. (2005). *Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS). Theoretische und methodische Grundlagen, Konstruktvalidität und psychometrische Eigenschaften bei der Beschreibung intra- und interindividueller Unterschiede. Forschungsberichte aus dem Projekt: <<Qualität des Erlebens in Arbeit und Freizeit>> (Nr. 6).* Zürich, Switzerland: Fachrichtung Angewandte Psychologie des Psychologischen Instituts der Universität.

Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015a). Introducing the familiarity mechanism: A unified explanatory approach for the personalization effect and the examination of youth slang in multimedia learning. *Computers in Human Behavior, 43,* 129–138. http://dx.doi.org/10.1016/j.chb.2014.10.052

Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015b). Mind your Ps and Qs! How polite instructions affect learning with multimedia. *Computers in Human Behavior, 51,* 546–555. http://dx.doi.org/10.1016/j.chb.2015.05.025

Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction, 44,* 65–73. http://dx.doi.org/10.1016/j.learninstruc.2016.03.002

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–70). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511816819.005

Schüler, A., Arndt, J., & Scheiter, K. (2015). Processing multimedia material: Does integration of text and pictures result in a single or two interconnected mental representations? *Learning and Instruction, 35,* 62–72. http://dx.doi.org/10.1016/j.learninstruc.2014.09.005

Schworm, S., & Stiller, K. D. (2012). Does personalization matter? The role of social cues in instructional explanations. *Intelligent Decision Technologies, 6,* 105–111. http://dx.doi.org/10.3233/IDT-2012-0127

Sherman, G. D., & Haidt, J. (2011). Cuteness and disgust: The humanizing and dehumanizing effects of emotion. *Emotion Review, 3,* 245–251. http://dx.doi.org/10.1177/1754073911402396

Sung, E., & Mayer, R. E. (2012). When graphics improve liking but not learning from online lessons. *Computers in Human Behavior, 28,* 1618–1625. http://dx.doi.org/10.1016/j.chb.2012.03.026

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4,* 295–312. http://dx.doi.org/10.1016/0959-4752(94)90003-5

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22,* 123–138. http://dx.doi.org/10.1007/s10648-010-9128-5

Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychological Review, 102,* 671–683. http://dx.doi.org/10.1037/0033-295X.102.4.671

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104,* 485–498. http://dx.doi.org/10.1037/a0026609

Vansteenkiste, M., Simons, J., Lens, W., Soenens, B., Matos, L., & Lacante, M. (2004). Less is sometimes more: Goal content matters. *Journal of Educational Psychology, 96,* 755–764. http://dx.doi.org/10.1037/0022-0663.96.4.755

Wang, Y., & Crooks, S. M. (2015). Does combining the embodiment and personalization principles of multimedia learning affect learning the culture of a foreign language? *Journal of Educational Multimedia and Hypermedia, 24,* 161–177.

Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences, 14,* 383–388.

Waytz, A., Klein, N., & Epley, N. (2013). Imagining other minds: Anthropomorphism is hair-triggered but not hare-brained. In M. Taylor (Ed.), *The Oxford handbook of the development of imagination* (pp. 272–287). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780195395761.013.0018

Woo, H. L. (2009). Designing multimedia learning environments using animated pedagogical agents: Factors and issues. *Journal of Computer Assisted Learning, 25,* 203–218.

Zander, S., Reichelt, M., Wetzel, S., Kämmerer, F., & Bertel, S. (2015). Does personalisation promote learners' attention? An eye-tracking study. *Frontline Learning Research, 3,* 1–13.

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics, 7,* 347–360. http://dx.doi.org/10.1007/s12369-014-0267-6

Appendix A

Excerpt of the Learning Content

The learning texts deals with the subject of artificial intelligence (AI). Based on Elaine Rich, AI is defined by a description of how digital information processing is compared to processing of humans. After defining basic concepts, the Turing test is described as a paradigm which can determine whether a computer can be seen as intelligent being or an automatic robot. The second section contains information about AI and perception processes. Therefore, the human brain and sensory organs are described as basis for robot designs. Several examples are outlined which describe to what extend AI deals with the problem of the complex processes behind perception and interaction with the environment. The third section deals with AI and awareness. The human awareness is described and approaches how to generate awareness in a computer system are shown. Thus, the implementation of neuronal networks, potential problems and opportunities for computer technology are outlined. The last section deals with AI in searching processes. Problem solving can be defined as searching process in the state of construction. Furthermore, the general problem solver, which was created from Newell and Simon, is described in detail.

(Appendices continue)

Appendix B

The Adapted Cognitive Load Questionnaire From Leppink et al. (2014) in the German Version Used in Experiment 2 and a Subsequently Translated English Version

Die zehn folgenden Fragen beziehen sich auf die Lernumgebung, die Sie/du vorher bearbeitet haben/hast. Bitte lesen Sie/lies jede Frage aufmerksam durch und antworten Sie/antworte mit Hilfe der Skala von 0 bis 10, wobei 0 “gar nicht zutreffend” und 10 “voll und ganz zutreffend” bedeutet.

0 1 2 3 4 5 6 7 8 9 10

[1] Die Themen/Das Thema in der Lernumgebung waren/war kompliziert.

[2] Die Lernumgebung beinhaltete Sachverhalte, die ich als komplex empfunden habe.

[3] Die Lernumgebung beinhaltete Konzepte und Definitionen, die ich als kompliziert empfunden habe.

[4] Die Instruktionen und/oder Erklärungen in der Lernumgebung waren sehr unklar.

[5] Die Instruktionen und/oder Erklärungen waren nicht hilfreich für das lernen.

[6] Die Instruktionen und/oder Erklärungen wurden sprachlich ungenau beschrieben.

[7] Die Lernumgebung hat mein Verständnis für das bearbeitete Thema/die bearbeiteten Themen verbessert.

[8] Die Lernumgebung hat mein Wissen und Verständnis zu dem Thema “Künstliche Intelligenz” verbessert.

[9] Die Lernumgebung hat mein Verständnis für die einzelnen bearbeiteten Sachverhalte verbessert.

[10] Die Lernumgebung hat mein Verständnis für die bearbeiteten Konzepte und Definitionen verbessert.

All of the following 10 questions refer to the previously handled learning environment. Please take your time to read each of the questions carefully and respond to each of the questions on the presented scale from 0 to 10, in which ‘0’ indicates not at all the case and ‘10’ indicates completely the case:

0 1 2 3 4 5 6 7 8 9 10

[1] The topic/topics covered in the learning environment was/were very complex.

[2] The learning environment covered matters that I perceived as very complex.

[3] The learning environment covered concepts and definitions that I perceived as very complex.

[4] The instructions and/or explanations within the learning environment were very unclear.

[5] The instructions and/or explanations were, in terms of learning, very ineffective.

[6] The instructions and/or explanations were full of unclear language.

[7] The learning environment really enhanced my understanding of the topic(s) covered.

[8] The learning environment really enhanced my knowledge and understanding of the topic “Artificial Intelligence.”

[9] The learning environment really enhanced my understanding of the matters covered.

[10] The learning environment really enhanced my understanding of the covered concepts and definitions.

Received December 8, 2016
Revision received March 24, 2017
Accepted March 26, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 233–249

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000209

How Affective Charge and Text–Picture Connectedness Moderate the Impact of Decorative Pictures on Multimedia Learning

Sascha Schneider, Jonathan Dyrna, Luis Meier, Maik Beege, and Günter Daniel Rey
Chemnitz University of Technology

Decorative pictures, which make a learning text aesthetically appealing rather than provide information, have been predominantly found to impair learning by an increase of learning-irrelevant cognitive processes. Recent research, however, indicates that this effect is moderated by various factors. On the basis of cognitive–affective theories and studies, the affective charge and the degree of text–picture connectedness (i.e., the semantic relation of text and pictures) of decorative pictures reveal possible boundary conditions. To examine these design features and compare them with a group without pictures, 3 experiments ($N_1 = 108$; $N_2 = 86$; $N_3 = 162$) with secondary school (Experiments 1 and 3) or university (Experiment 2) students were conducted. For this, decorative pictures consistent with those in instructional texts about South Korea (Experiments 1 and 2) or the human body (Experiment 3), were tested in a 2 (positively vs. negatively charged) × 2 (weakly vs. strongly connected to the text) between-subjects design with an additional control group. Learning performance, affective responses, and cognitive processes were measured. Results show that students with either positive or strongly connected pictures outperformed students with negative or weakly connected pictures. In comparison with the control group, strongly connected positive pictures enhanced learning and weakly connected negative pictures impaired learning. Although negative pictures were shown to increase task-irrelevant thoughts and extraneous cognitive load, weakly connected pictures increased the perception of intrinsic cognitive load.

Educational Impact and Implications Statement

This study reveals that incorporating decorative pictures within multimedia materials is beneficial for learning when pictures are positive or strongly connected to the content of the text rather than negative or weakly connected. This is explained by an increase of task-irrelevant thoughts for negative pictures and an increase of perceived task complexity for weakly connected pictures. In addition, the inclusion of strongly connected, positive pictures support learning, whereas negative, weakly connected pictures inhibit learning in contrast to a text-only condition. In conclusion, decorative pictures might be used to enrich learning material if boundary conditions like the degree of connectedness or affective charge are taken into account.

Keywords: decorative pictures, affective charge, text–picture relation, boundary conditions, multimedia learning

How do decorative pictures interspersed into multimedia learning materials influence learning? This question is of crucial empirical and practical importance because instructional material designers tend to add not only information but also decorative photographs and illustrations to learning materials (Pozzer & Roth, 2003). Consequently, a notable body of research was conducted to answer the question of whether the inclusion of decorative pictures is beneficial or detrimental to learning processes (e.g., Danielson, Schwartz, & Lippmann, 2015; Schneider, Nebel, & Rey, 2016; Sung & Mayer, 2012). Early findings suggest that such pictures impair learning because, for example, they interrupt schema construction or distract learners' attention (e.g., Harp & Mayer, 1998). More recent studies, however, reveal learning-enhancing effects of decorative pictures when used as metacognitive support (e.g., Danielson et al., 2015). The present study aimed at providing evidence for potential boundary conditions of interspersed decorative pictures so as to answer the initial question.

This article was published Online First June 5, 2017.

Sascha Schneider, Jonathan Dyrna, Luis Meier, Maik Beege, and Günter Daniel Rey, E-Learning and New Media, Faculty of Humanities, Chemnitz University of Technology.

Correspondence concerning this article should be addressed to Sascha Schneider, E-Learning and New Media, Faculty of Humanities, Technische Universität Chemnitz, Straße der Nationen 12, 09111 Chemnitz, Germany. E-mail: sascha.schneider@phil.tu-chemnitz.de

Theoretical Framework

Cognition and Affect in Multimedia Learning

In multimedia learning research, cognitive processes are often explained on the basis of cognitive load theory (CLT; Sweller, 2010). CLT assumes that information is consciously processed in human working memory, which is of limited capacity. When this

limit is exceeded, working memory becomes overloaded and learning (i.e., schema construction and modification in long-term memory) might be inhibited. Recent versions of the theory distinguish between two distinct types of cognitive load, namely intrinsic and extraneous cognitive load (Kalyuga & Singh, 2016). *Intrinsic cognitive load* deals with any cognitive activities that are related to the concurrent processing of interacting information elements in working memory and their integration within existing knowledge in accordance with a specific learning objective. Earlier versions of this theory subsumed all integration processes in an additional load called *germane cognitive load* (Sweller, 2010). *Extraneous cognitive load*, in contrast, refers to any cognitive processes that are not primarily necessary for learning and is imposed by cognitive activities that result from the way an instructional message is organized or presented (i.e., instructional design; Sweller, 2010). According to CLT, an instructional message should be designed in a way that (a) optimizes intrinsic load (e.g., by selecting elements that match the learner's expertise), (b) minimizes extraneous cognitive load (e.g., by excluding unnecessary information), and (c) motivates learners to allocate unassigned working memory resources to learning-relevant processes (Sweller, 2010). However, the CLT's explanatory power is limited to the extent that responses of affect are not taken into account (Huk & Ludwigs, 2009), although affect is naturally interconnected with cognitive processes (Gläser-Zikuda, Fuß, Laukenmann, Metz, & Randler, 2005; Levine & Pizarro, 2004; Pekrun, 2006; Plass & Kaplan, 2015; Schwarz, 2000).

Suggestions to consider affective influences on learning and instruction are met by integrated cognitive–affective frameworks such as the cognitive–affective theory of learning with media (CATLM; Moreno & Mayer, 2007) or the integrated cognitive–affective model of multimedia learning (ICALM; Plass & Kaplan, 2015). CATLM, for example, proposes that learners' cognitive processing (i.e., selection, organization, and integration) of information in a multimedia message is influenced by affective factors (Moreno & Mayer, 2007). According to ICALM, the way learners process information is impacted by their responses to (core) affect, which is, for instance, evoked by the message's specific design (Plass & Kaplan, 2015). Core affect constitutes a neurophysiological state which is consciously accessible as a simple, nonreflective feeling that is a combination of two dimensions, namely arousal (activation–deactivation) and valence (pleasure–displeasure; Russell, 2003). Although frequently considered independent from each other, current research points to the idea of arousal as a weak but consistent V-shaped function of valence in subjective experience (Kuppens, Tuerlinckx, Russell, & Barrett, 2013). Because the encountered nomothetic relation is overshadowed by an observation of large individual differences, we instead refer to the (empirically) more robust dimensions of positive affect (also named positive activation) and negative affect (also named negative activation; Tellegen, Watson, & Clark, 1999). Positive affect comprises positive states of high activation (e.g., excitement) and negative states of low activation (e.g., boredom). Analogously, negative affect ranges from highly activating negative states (e.g., anxiety) to positive states with a low degree of activation (e.g., relaxation). When attributed to a learning object, core affect begins an emotional episode (Russell, 2003; Shuman & Scherer, 2014) that comprises multiple components (e.g., phenomenological, expressive, and motivational components; Roseman, 2011; Shuman & Scherer, 2014). This episode is of stronger intensity but shorter duration compared with moods (Russell, 2003; Shuman & Scherer, 2014). For instance, looking at an aesthetically pleasing (decorative) picture that depicts a baby lynx might make a learner experience feelings of pleasure (phenomenological component), put a smile on his or her face (expressive component), and encourage him or her to read the text (motivational component). As a result, experienced emotions might impact cognitive processing, such as the allocation of cognitive resources (Huk & Ludwigs, 2009) and the selection of information. In addition, emotions are organized and integrated into affective–cognitive mental representations (i.e., schemata) of the subject matter together with the mental representation of a learning material (Plass & Kaplan, 2015). All of the cognitive–affective theories hypothesize, however, that affective influences of single design features should be taken into account.

Decorative Pictures in Multimedia Learning

Pictures are external (knowledge) representations that, for instance, vary in terms of colorfulness (e.g., achromatic vs. colorful; Um, Plass, Hayward, & Homer, 2012) or degree of abstraction (e.g., grounded vs. idealized; Belenky & Schalk, 2014). In educational research, a multitude of taxonomies attempt to categorize pictures regarding their role in text processing (e.g., Lee & Nelson, 2004; Levie & Lentz, 1982; Marsh & White, 2003; Mayer, 1993), thereby distinguishing between four (i.e., representation, organization, explanation, and decoration; Mayer, 1993) up to 49 (Marsh & White, 2003) distinct functions. The majority of such taxonomies identify decorative pictures as external representations that express either little or no relation to the text (e.g., Levin, 1981). In contrast to most other types of pictures that are relevant to the instructional goal (i.e., instructive pictures; e.g., Sung & Mayer, 2012), decorative pictures are frequently considered irrelevant to the essential material but might enable aesthetic experience (e.g., Takahashi, 1995). Beyond theory-based taxonomies, pictures commonly serve more than one function at the same time (Danielson et al., 2015). Consequently, the aforementioned functions (information and decoration) do not inevitably exclude each other but can rather be considered as two orthogonal dimensions (Lenzner, Schnotz, & Müller, 2013). On the basis of the proposed dimensional approach, pictures that aim to make learning material aesthetically appealing rather than provide information can be defined as *decorative pictures*. For example, a portrait of a baby lynx, as an example for a native Korean animal, might be added to an instructional text about Korean animals to evoke an aesthetic appeal but might also provide a small amount of learning-relevant information (i.e., the lynx is an example for a native Korean animal).

Because of their limited relation to the instructional objective, decorative pictures are frequently considered as extraneous materials that impede learning by overloading the cognitive capacities of learners and, therefore, should be excluded from a multimedia message (i.e., *coherence principle;* Mayer & Fiorella, 2014). This adverse effect, which is even enhanced when elements of the pictures are of a high interest (i.e., *seductive detail effect*; e.g., Mayer, Griffith, Jurkowitz, & Rothman, 2008; Sung & Mayer, 2012), received multiple empirical (e.g., Harp & Mayer, 1998) as well as meta-analytical support (Rey, 2012). However, findings in

this field are not consistent—with studies showing no inhibiting (e.g., Chang & Choi, 2014; Sitzmann & Johnson, 2014) or even partially beneficial influences (Chen & Latham, 2014; Danielson et al., 2015; Lenzner et al., 2013; Wang & Crooks, 2015). These inconsistent findings lead to the assumption that boundary conditions moderate the influence of decorative pictures.

By reviewing current empirical findings in this field of research, we identified three potential moderators. First, the effect is influenced by the general setting of the learning or retrieval phase, for instance the inclusion or exclusion of a time limit (Rey, 2012). Second, learners' individual characteristics, for example their level of prior knowledge (Magner, Schwonke, Aleven, Popescu, & Renkl, 2014) and working memory capacity (Sanchez & Wiley, 2006), likely moderate the impact of decorative pictures. Third, results might be influenced by the design of the multimedia material. On the one hand, modifying the pictures' individual design might aid learning by enabling further beneficial functions. For instance, conducive decorative pictures (Schneider et al., 2016), which are defined as learning-enhancing decorative pictures, might provide metacognitive support (Chen & Latham, 2014), enhance learners' emotional states, or serve as metaphorical aids (Danielson et al., 2015). On the other hand, design properties of the entire multimedia message, such as the modality of the text (Park, Moreno, Seufert, & Brünken, 2011; Park, Flowerday, & Brünken, 2015) or the relation between the interacting elements, such as their arrangement and the degree of text–picture relation, should also be taken into account as boundary conditions. In the present study, we focus on the impact of such multimedia design facets. As suggested by Schneider et al. (2016), we aim to investigate the influence of two potential boundary conditions, namely the affective charge of decorative pictures and their degree of relation to the alongside presented text.

Affective Charge of Decorative Pictures

Besides cognitive aids, pictures are frequently suggested to fulfil affective functions in text processing (e.g., Levie & Lentz, 1982; Marsh & White, 2003). For instance, pictures might evoke a multitude of affective states, somewhat depending on the beholder's individual perception (e.g., Bradley, Codispoti, Cuthbert, & Lang, 2001). As suggested by Schneider et al. (2016), appealing (i.e., positive emotional) decorative pictures might contribute to an aesthetically pleasing design of multimedia learning materials. Positive affective (hereafter referred to as *positive*) states elicited through decorative pictures were shown to foster learning outcomes compared with both neutral (Park, Knörzer, Plass, & Brünken, 2015; Plass, Heidig, Hayward, Homer, & Um, 2014; Um et al., 2012) and negative affective (hereafter referred to as *negative*) states (Heidig, Müller, & Reichelt, 2015). In a pilot study (Schneider et al., 2016), students who processed learning materials on cell division with interspersed positive emotional decorative pictures outperformed students with negative pictures, whereas learning performance was mediated by higher amounts of perceived pleasure. Despite first crucial insights, these findings remain limited because the influence of elicited emotional states on concurrent cognitive processes was not furtherly investigated and no control group (without decorative pictures) was included. This is of particular importance because an emotional overload might exceed the learners' cognitive capacities and lead their attention away from the learning activities (Plass & Kaplan, 2015). Consequently, a text-only condition might outperform both the positive and the negative pictures group because elicited emotions might impose extraneous cognitive load. As an alternative approach, we state that mainly negative emotions induced by decorative pictures might evoke task-irrelevant thoughts (Pekrun, Goetz, Titz, & Perry, 2002) that contribute to extraneous cognitive load. For instance, a decorative picture displaying caged Korean Jindo dogs might evoke feelings of frustration that is accompanied by extraneous thoughts about how cruel humans can be. In contrast, positive emotions, such as enjoyment of learning, direct attention toward the learning task and allow a full use of cognitive resources to achieve the instructional objectives (Huk & Ludwigs, 2009; Pekrun et al., 2002). Moreover, elicited positive emotions might foster cognitive processing due to more divergent and creative thinking compared with negative emotional episodes (e.g., Isen, Daubman, & Nowicki, 1987; Nadler, Rabi, & Minda, 2010). This assumption might serve as an explanation for large detrimental effects of extraneous materials that particularly occur when such elements contain information that might be linked to negative emotions (Schneider et al., 2016), for instance life-threatening consequences of lightning (Harp & Mayer, 1998) or mobile phone users being involved in car wrecks (Chang & Choi, 2014).

The Relation of Text and Decorative Pictures

Recent evidence suggests that learners always actively construct relations between a learning text and accompanied pictures (Danielson et al., 2015; Danielson & Sinatra, 2016) to build integrated mental models (Schnotz, 2014), irrespective of the pictures' relevance for the learning goal. To what extent such a relation is established might be strongly influenced by two facets of multimedia design. On the one hand, the layout of a learning material (i.e., the arrangement of pictures and text) was found to impact its comprehensibility. For example, when pictures are positioned too distant from related text sources, learners are required to split their attention between text and picture so that the construction of integrated models is inhibited (Chandler & Sweller, 1992). On the other hand, the pictures' content might encourage or impede learners to (semantically) relate the depicted information to the text. In this context, decorative pictures might vary in their semantic connectedness to the text, ranging from weakly to strongly connected. Whereas, for example, a decorative picture of a baby lynx is strongly connected to a zoological text about endangered species, the same picture is rather weakly connected to a text about people in South Korea. Considering both cognitive and affective effects, we suggest four hypotheses that might explain why a strong semantic text–picture connectedness contributes to an increased learning performance.

According to the advantageous imagination hypothesis, when learners are encouraged to imagine a concept as they study, they learn better in contrast to when they are instructed to study only (i.e., the imagination effect; Leahy & Sweller, 2008). This effect can be explained by an increase of transferring knowledge from working memory to long-term memory. Decorative pictures might function as an implicit instruction to imagine a concept and thereby encourage learners to invest more cognitive resources into the construction of learning-relevant schemata. In case the content of the representation is strongly connected to the to-be-learned

information, the construction of appropriate schemata is fostered. In contrast, schema construction is hampered by weakly connected picture because a possible imagination leads to learning-irrelevant schemata which hampers learning.

According to the high cohesion hypothesis, when events (in this case, text and pictures) are too vague or the connection of the elements is hardly understandable (i.e., low cohesion), learners are required to generate more inferences and use more resources to form a coherent representation (Linderholm et al., 2000; Sukalla, Bilandzic, Bolls, & Busselle, 2015). Because more resources are associated with a higher element interactivity (Kalyuga & Singh, 2016), weakly connected decorative pictures lead to an increased perception of intrinsic cognitive load and reduced resources for learning. This interfering impact can be resolved by more precise connections of elements provided by strongly connected decorative pictures (i.e., high cohesion).

According to the fluency hypothesis, a high degree of text–picture connectedness might also increase users' *information processing fluency* (Van Rompay, De Vries, & Van Venrooij, 2010), which is the subjective experience of ease with which individuals process information (Alter & Oppenheimer, 2009). The *hedonic fluency model* (Winkielman, Schwarz, Fazendeiro, & Reber, 2003) suggests that high processing fluency evokes a genuine affective reaction that is hedonically positive. Evidence for this proposal is drawn from a study using psychophysiological affective measures (Winkielman & Cacioppo, 2001). Regarding learning materials, we assume that a high degree of connectedness between texts and pictures might increase learners' positive affect and foster learning.

According to the flow hypothesis, when decorative pictures provide scant learning-relevant information, the text–picture relation might not be obvious to the learner. Subsequently, such impasses lead to momentary confusion (D'Mello & Graesser, 2012). When decorative pictures are strongly connected to the topic, learners are able to resolve such confusion more readily through faster identification of the text–picture relation and, thus, can continue to attain the learning objective (i.e., a flow situation; D'Mello & Graesser, 2012). In contrast, when decorative pictures are more weakly connected, increased confusion might lead to feelings of frustration (i.e., increased negative affect), which result in lower learning performance.

Research Questions

On the basis of theoretical explanations of previous empirical results (e.g., D'Mello & Graesser, 2012; Leahy & Sweller, 2008; Schneider et al., 2016; Sukalla et al., 2015; Winkielman & Cacioppo, 2001), we propose that two additional design features—namely the nature of affective charge and the text–picture connectedness of decorative pictures—impact the learners' cognitive processing and learning outcomes. In this regard, positive affect or strong text–picture connectedness are expected to promote learning in contrast to negative affect or weakly connected decorative pictures.

> *Question 1:* How do affective charge and the degree of text–picture connectedness of decorative pictures interspersed into multimedia materials influence cognitive processing and learning outcomes?

To reveal answers to the first research question regarding whether specific configurations of decorative pictures are conducive or detrimental for learning in contrast to the exclusion of decorative pictures, we need to compare the learning results of students with decorative pictures and in a text-only condition. These comparisons should further a discussion about the use of such pictures in multimedia learning design. In this context, we assume that positive and strongly connected pictures foster learning, whereas negative and weakly connected pictures hinder learning. More specifically, we addressed the following research question:

> *Question 2:* Are there decorative pictures that can be identified as learning-enhancing (conducive pictures) or learning-inhibiting (detrimental pictures)?

> *Question 3:* Do students with conducive or detrimental pictures reveal differences in their assessed cognitive processing?

Experiment 1: Prestudy

Because this study aims to apply the more robust factor structure of positive and negative affect (Tellegen et al., 1999), prevalidated picture databases such as the International Affective Picture System (Lang, Bradley, & Cuthbert, 2008) were not appropriate. In line with comparable multimedia design studies (e.g., Magner et al., 2014), we conducted our own validation to identify affective decorative pictures that are appropriate for the South Korea learning topic. This prestudy aimed to determine which decorative pictures induced the most positive or negative affect.

Method

Participants and design. The study was carried out with a one-factor within-subjects design, testing two states of perceived affect in pictures: positive affect and negative affect. In total, 66 participants (60.6% female) voluntarily took part in the experiment. The mean age was 27.6 years (SD = 8.4). All participants reported their native language to be German. Furthermore, 10.6% of the participants were either vegans or vegetarians.

Materials. Three categories of motifs that might serve as appropriate elicitors for both positive and negative affect, irrespective of the individuals' gender, were identified on the basis of previous research on emotional pictures (e.g., Lang et al., 2008). First, human faces portraying emotions have been successfully applied in a wide range of studies to induce affect (e.g., Schneider et al., 2016). Therefore, two native Korean models (one female and one male) were instructed to show specific positive (e.g., excitement, pride) and negative (e.g., anger, disgust) activating emotions.

Second, animals might elicit emotions, depending on their situational state (e.g., Lang et al., 2008). Babies, in general, were shown to induce positive affect (e.g., excitement). In contrast, animals that are either perceived as threatening (e.g., spiders) or endangered (e.g., caged Jindo dogs) mainly evoke negative affect (e.g., anxiety, anger). Consequently, pictures of both native Korean baby animals and endangered animals were selected from web resources.

Third, food is appropriate to induce both positive and negative emotions (e.g., Lang et al., 2008). Whereas an appealing presen-

tation of food is associated with states of positive affect (e.g., excitement), disgusting motifs such as a spoiled meat most likely evoke negative affect (e.g., disgust). Accordingly, we selected pictures of Korean dishes of assumed high and low appeal (considered for Western societies). With the help of a rater, who was familiar with South Korea and its culture, 87 pictures in total were selected to use for the prestudy, of which 20 portrayed humans of Korean origin (10 for each affective dimension), 22 depicted native Korean animals (22 positive and 21 negative), and 24 showed typical Korean dishes (10 positive and 14 negative).

Measures. The affective state elicited by the pictures was assessed with the German positive affect and negative affect scales of the Positive Affect, Negative Affect and VAlence short scales (PANAVA-SS, Schallberger, 2005). To ensure brevity and avoid high dropout rates, two items of each scale were selected (i.e., "excited–bored" and "highly motivated–weary" for positive affect; $\alpha = .91$; "stressed–relaxed" and "nervous–calm" for negative affect; $\alpha = .90$ [translated into English for the present article]). The items were rated on 7-point scales. Demographic data of learners' age, sex, and native language (German or other) were collected. Because some of the raters' appraisals of the depicted motifs, for instance the displayed food, might be strongly influenced by individual characteristics such as special eating habits (e.g., vegetarianism), the participants' diet was additionally assessed. Intraclass correlation shows that interrater reliability for the mean valence scores, ICC $(2, k) = .995$, $F(1, 65) = 216.25$, $p < .001$, can be assessed as almost perfect (Landis & Koch, 1977).

Procedure. The prestudy was conducted online. First, all demographic characteristics were obtained. Second, a short introduction and one example question were given. Subsequently, each picture was shown on a separate web page. The pictures had to be assessed concerning the question, "How did you feel while looking at the picture?", by means of the displayed scales for positive and negative affect. In total, each participant was shown between 45 and 87 pictures in a random sequence. To avoid fatigue effects, the participants were allowed to cease the rating of pictures at any time. Overall, 83% of the participants rated all items.

Results

To analyze differences between ratings of positive and negative pictures, two mixed-factors univariate analyses of variance were conducted with the mean positive and negative affect scores for each category of pictures as the within-subject factor and age, gender, and diet as the between-subjects factors, with age used as continuous variable.

Regarding positive affect, results showed a significant main effect, (Wilk's $\Lambda = 0.22$), $F(3, 11) = 13.22$, $p = .001$, $\eta_p^2 = .78$. Positive pictures ($M = 4.64$, $SD = 0.96$) were assessed as more positive than were negative pictures ($M = 3.31$, $SD = 0.64$). The interaction terms of positive affect and between-subjects factors were not significant ($p = .39, .87$). Follow-up tests showed significant differences in terms of positive affect for all three distinct categories of pictures, namely humans ($p = .002$, $\eta_p^2 = .55$), fauna ($p = .012$, $\eta_p^2 = .40$), and food ($p < .001$, $\eta_p^2 = .77$).

Regarding negative affect, results revealed a significant main effect, (Wilk's $\Lambda = 0.07$), $F(3, 11) = 50.30$, $p < .001$, $\eta_p^2 = .93$. Negative pictures ($M = 4.14$, $SD = 1.21$) were perceived as more negative than were positive pictures ($M = 2.32$, $SD = 0.69$). The interaction of negative affect and between-subjects factors was not significant ($p = .16, .82$). Follow-up tests showed significant differences in terms of negative affect for all three distinct categories of pictures, namely humans ($p = .007$, $\eta_p^2 = .44$), fauna ($p < .001$, $\eta_p^2 = .68$), and food ($p < .001$, $\eta_p^2 = .92$). These results revealed that the preselected positive and negative pictures evoked the respective affective charge. Consequently, among the pictures with the highest scores in terms of corresponding affective dimension, either two (fauna) or three (food and humans) pictures with different motifs were selected (see Figure 1).

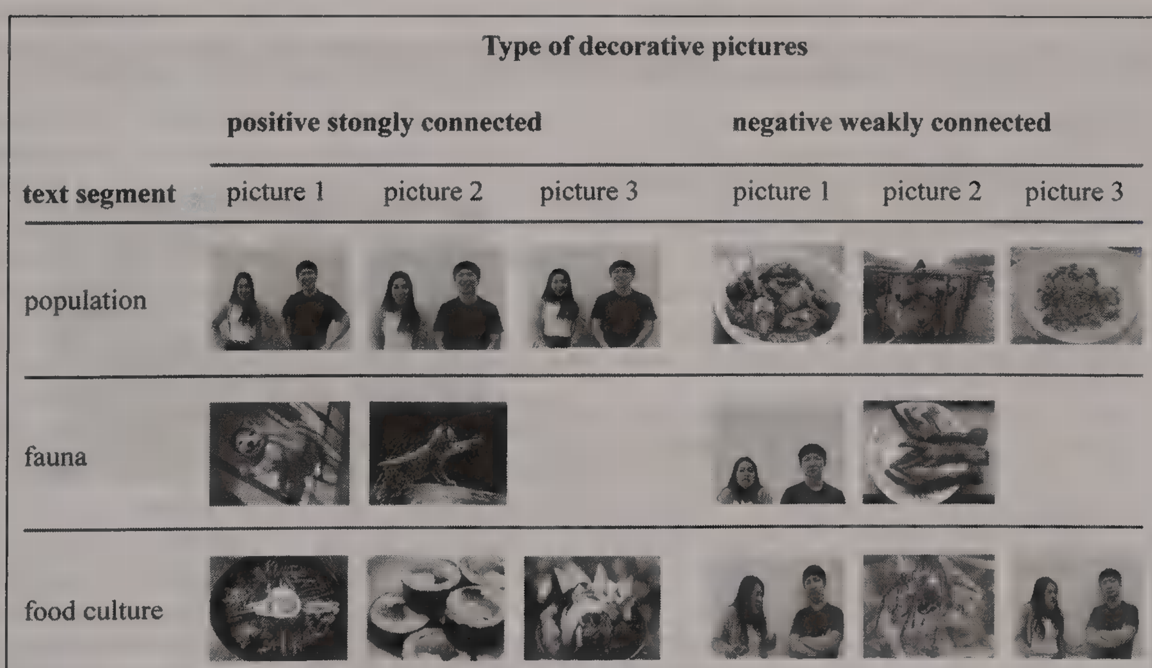

Figure 1. Experimental pictures used in Experiment 1 and 2 together with a learning text about South Korea. See the online article for the color version of this figure.

Experiment 1

Method

Participants and design. Overall, 108 secondary school students (60.2% female) from Germany took part in this experiment. The mean age was 17.22 years (SD = 0.89). Students were either in Grade 11 (51.9%), 12 (34.3%), or 13 (13.9%) and majored in economy (42.6%), design and media (38.0%), or health and social issues (19.4%). All students reported their native language to be German. Mean prior knowledge (further described in the following Learning materials subsection) was 0.33 out of 3 points (SD = 0.58).

This experiment aimed at varying the affective charge and the degree of text–picture connectedness (hereinafter called *connectedness*) of decorative pictures. Students were assigned to one of the four experimental groups of a 2 × 2 between-subjects design or the additional control group without decorative pictures (C group; N = 19) via random assignment by a computer algorithm. In conclusion, students received materials with either positive and strongly connected pictures (PS group; N = 27), negative and strongly connected pictures (NS group; N = 17), positive and weakly connected pictures (PW group; N = 19), or negative and weakly connected pictures groups (NW group; N = 27).

Learning materials. The learning material consisted of 1,631 words and either zero (C) or eight decorative pictures, depending on the experimental condition, which were chosen on the basis of the prestudy. All learning materials were displayed on a computer screen. The learning text consisted of facts about South Korea and was separated into three segments: "Population," "Fauna," and "Food culture." The text was appropriate for learners at the end of secondary education. Each segment was separated into chapters, whereby each chapter was displayed on a separate web page. The length of these chapters were chosen so that they did not exceed the size of a web page (for an example web page, see Figure 2). For all groups, except for the control group, each chapter was displayed together with one decorative picture. The included pictures differed according to the experimental variation; that is, positive pictures were shown in positive emotional groups, and negative pictures were shown in negative groups. In the cases of groups with weakly connected pictures, the sequence of pictures was mixed in a predefined order (as displayed in Figure 1) so that no segment contained pictures that could be easily connected to the corresponding text. Readers started on a menu page that consisted of the heading of the learning text ("South Korea") and three buttons that led to the chapter segments. Users were allowed to navigate independently through all web pages. The last pages of the chapter segments included a button leading back to the menu page. If participants clicked through all pages of one segment, a checkmark was displayed on the menu page beside the corresponding segment button. In addition, an exit button was displayed on the menu page, which led to the first of two versions of a finishing page that asked participants if they wanted to continue on to the next part of the experiment or if they wanted to go back to the learning environment. For each possibility, a separate button was displayed. In addition, all pages were headlined with a time bar that counted down from 17 min to 0 min. This maximum amount of time was specified on the mean readings times of five nonexpert pretest readers (M = 15.1 min, SD = 1.1) who had not read the text before. If the reading time expired, participants were automatically directed to a second version of a finishing page where no back button existed. However, an analysis of the HTML protocols showed that not one of the participants was directed to this page (average learning time: M = 13.92 min, SD = 2.11).

Learning measures. A prior knowledge questionnaire was created. This questionnaire consisted of three questions in an open-answer format, which aimed at measuring the level of knowledge of learning-material relevant information. The questions were as follows: "What is the capital city of South Korea?", "What is the official currency of South Korea?", "What are typical dishes of South Korea?" Questions 1 and 2 were rewarded with one point

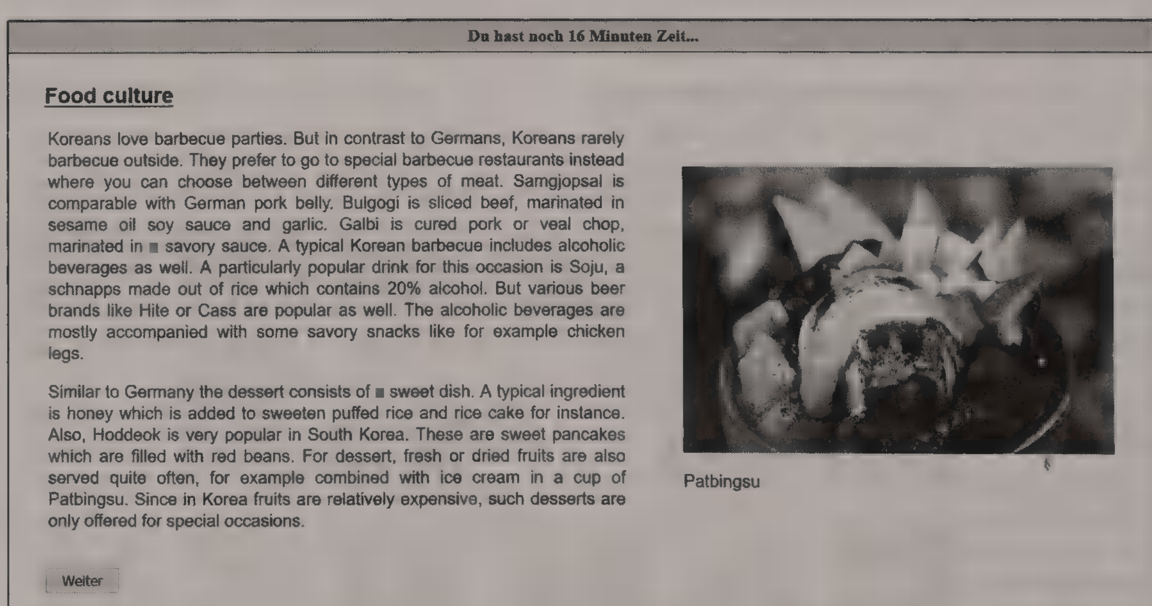

Figure 2. Example website of the learning environment for the segment "food culture" in Experiment 1 and 2 and the experimental group with positive and strongly text-connected decorative pictures. The text was translated from German into English. See the online article for the color version of this figure.

each if the correct answer was given. Because no discrepancies occurred between two raters, interrater reliability was perfect. Concerning the third question, one point was assigned for each correct answer. A maximum of five points could be reached. The sum of points of all prior knowledge questions was used as the prior knowledge score. For this, a maximum of seven points could be achieved with all three prior-knowledge questions ($\alpha = .70$).

A second task was designed to measure retention knowledge. This questionnaire consisted of four questions in an open-answer format and four multiple-choice questions with one correct answer. The open-answer questions (e.g., "What is the national dish of South Korea?") aimed at remembering terms introduced in the learning materials. Intraclass correlation of three trained raters showed that interrater reliability, ICC (2, k) = [.98, 1.00], $F(107, 107) = [63.51, 101.19]$, $p < .001$, can be assessed as almost perfect. To answer single-choice questions, learners had to choose the correct answer out of four possible answers according to a question on facts that were in the learning material. For example, students were given the answer possibilities "Gi," "Heung," "Hue," and "Jeong" to answer the question, "Which term does not characterize the Korean philosophy?" If the correct answer was given, learners were rewarded with one point per correct answer. A maximum of eight points could be reached within this knowledge category ($\alpha = .66$).

A third task was created to measure transfer knowledge. This questionnaire consisted of one question in an open-answer format and seven multiple choice questions with one correct answer. The types of questions were equal in their format compared with the retention questions. However, while answering the transfer questions, learners were required to apply recently achieved knowledge about South Korea to new situations. One example question in this category was, "A Korean friend is visiting you in Germany. You are not sure what to prepare for dinner. Which of the following meals would most probably meet her Korean taste?" Participants were given the answer possibilities "sandwich," "pumpkin soup," "mixed grill," and "Vienna sausages and potato salad." For this question, students had to apply their knowledge about Koreans' favor for barbecue, which was mentioned in the text. Again, each question was rewarded with one point. Interrater reliability for the open-answer question, ICC (2, k) = .94, $F(107, 107) = 16.53$, $p < .001$, could be assessed as almost perfect. A maximum of eight points could be reached for all transfer questions ($\alpha = .73$).

Additional measures. Similar to the prestudy, the emotional dimensions of learners' positive affect ($\alpha = .70$) and negative affect ($\alpha = .76$) were measured by using the PANAVA-KS questionnaire (Schallberger, 2005). In this questionnaire, students had to rate eight items (four for positive affect and four for negative affect) regarding how they felt at the moment using antonymous adjectives (e.g., "bored–excited" for positive affect) on a 7-point scale. Because learners' emotional states were measured before and after the learning materials, difference scores were calculated after the experiment for the positive affect difference (PAD) and the negative affect difference (NAD).

To measure attentional diversion as an indicator for mental load, an adapted version of the nine-item test-irrelevant thinking scale ($\alpha = .87$) was included. This scale was taken from the Reactions to Tests questionnaire (Sarason, 1984) and is rated on a 7-point scale ranging from 1 (*I totally disagree*) to 7 (*I totally agree*). Because the test-irrelevant thinking scale is intended to address test-related situations, it had been adjusted by substituting the term *task* for *test* within each item to cover learning-related situations (e.g., "During the learning *task*, I thought about recent past events").

Because a time limit has been found to be a moderator of the seductive detail effect (Rey, 2012), data on perceived time adequacy were collected via two items (i.e., "There was enough time to completely read the learning text" and "I would have needed more time to completely read the learning text"), which were rated on a 7-point scale with the same anchors as were used in the task-irrelevant thinking questionnaire ($\alpha = .69$). In addition, another item for the manipulation check measuring connectedness was included (i.e., "The pictures fitted to the learning text") and rated on a 7-point scale with the same anchors.

Demographic data were collected through participants' responses to questions regarding age, sex, native language (German or other), grade level (11, 12, or 13), course profile (economy, social media, or health and social issues), special eating habits (none, vegetarian, or vegan), and arachnophobia (yes or no) were posed. The latter two categories of variables were measured because some decorative pictures displayed animal food or spiders. If a student were to have self-identified as having arachnophobia, he or she would have been excluded from further analyses; however, no participant self-identified thusly.

Procedure. In a classroom at the participating school, computers assigned to each workplace were prepared to display the starting page of the first questionnaire and a participants' number sheet, which allowed data to be combined across all parts of the experiment. Teachers escorted the students to this room before each lesson started and briefly explained the experimental situation. The experimenter introduced all tasks and parts of the experiment with a premade instruction form so as to increase comparability. After these instructions, students started with their computer environment. The experiment was separated into three parts. The first part comprised all questionnaires that collected data on learners' prior knowledge and current emotional states (PA and NA scales). At the end of this part, a computer program randomly assigned students to one of the five experimental conditions. The second part consisted of the learning materials. During the third part, demographic data were collected and all dependent variables or covariates were measured in the following order: the PA and NA, the connectedness, the retention and transfer, the task-irrelevant thinking, and the time adequacy scales. Students were allowed to ask technical questions only. All learners took between 35 min and 44 min to complete all three parts. Students were instructed to stay at their workplaces until everyone was ready. The experiment was conducted on 3 school days. Student group sizes differed between 15 and 21 students ($M = 17.2$, $SD = 1.1$).

Results

In the analysis of data, multivariate analyses of covariance (MANCOVAs) and univariate analyses of covariance (ANCOVAs) with affective charge and connectedness as between-subjects factors were conducted to assess differences between factor levels. In addition, subsequent covariance analyses for all dependent measures were performed with all five groups, whereby only differences to the control group were reported. Predefined test assumptions were reported only if significant violations occurred. There

were no significant differences between the five groups of the experiment in terms of age, grade, course of study, starting time, day of experiment, perception of time adequacy, and the baseline measurements of positive and negative affect (all $ps > .05$, except in terms of gender, $p = .008$). As a result, gender and prior knowledge as an important moderator of multimedia effects (Kalyuga, 2014) were included as additional covariates, whereby only significant influences of the covariate were reported. All descriptive results are shown in Table 1.

Manipulation check. First, a manipulation check was conducted to ensure that the experimental manipulation for connectedness succeeded. Thus, an ANCOVA was conducted with the manipulation check item on connectedness as dependent variable. A significant effect was found for connectedness, $F(1, 83) = 20.77$, $p < .001$, $\eta_p^2 = .20$, but not for affective charge, $F(1, 83) = 0.08$, $p > .05$, $\eta_p^2 < .01$, or the interaction, $F(1, 83) = 0.33$, $p > .05$, $\eta_p^2 < .01$.

Affective charge. To analyze if affective charge influenced the assessments of emotional states, a MANCOVA with PAD and NAD as dependent measures was conducted. Significant main effects were found for affective charge, (Wilk's $\Lambda = 0.74$), $F(2, 82) = 14.23$, $p < .001$, $\eta_p^2 = .26$, but not for connectedness, (Wilk's $\Lambda = 0.99$), $F(2, 82) = 0.20$, $p > .05$, $\eta_p^2 = .01$, or the interaction, (Wilk's $\Lambda = 0.99$), $F(2, 82) = 0.12$, $p > .05$, $\eta_p^2 < .01$. A follow-up ANCOVA for affective charge revealed significant effects for both the PAD, $F(1, 83) = 16.17$, $p < .001$, $\eta_p^2 = .16$, and the NAD scores, $F(1, 83) = 21.86$, $p < .001$, $\eta_p^2 = .21$. Students with strongly connected pictures reported higher scores of connectedness than did students with weakly connected pictures. In addition, students with negative pictures reported lower scores of PAD and higher scores of NAD than did students with positive pictures. In conclusion, the second manipulation check was fully confirmed.

Learning performance. To evaluate the influence of both manipulations on the retention and transfer scores, a MANCOVA was conducted with the retention and transfer scores as dependent measures. Significant main effects were found for affective charge, (Wilk's $\Lambda = 0.57$), $F(2, 82) = 31.05$, $p < .001$, $\eta_p^2 = .43$, and for connectedness, (Wilk's $\Lambda = 0.82$), $F(2, 82) = 9.14$, $p < .001$, $\eta_p^2 = .18$, but not for the interaction, (Wilk's $\Lambda = 0.99$), $F(2, 82) = 0.34$, $p > .05$, $\eta_p^2 = .01$.

Regarding retention, a follow-up ANCOVA revealed significant effects for both affective charge, $F(1, 83) = 43.84$, $p < .001$, $\eta_p^2 = .35$, and connectedness, $F(1, 83) = 10.86$, $p = .001$, $\eta_p^2 = .12$. Regarding transfer, significant effects were found for the affective charge of decorative pictures, $F(1, 83) = 36.72$, $p < .001$, $\eta_p^2 = .31$, and connectedness, $F(1, 83) = 16.21$, $p < .001$, $\eta_p^2 = .16$. In sum, positive or strongly connected decorative pictures fostered retention and transfer performance in contrast to negative and weakly connected pictures (see Table 1). Both affective charge and connectedness affected learning.

Differences from the control group. A subsequent MANCOVA was conducted with all five groups using the retention and transfer scores as dependent variables. A significant main effect was shown for group (Wilk's $\Lambda = 0.58$), $F(2, 82) = 7.97$, $p < .001$, $\eta_p^2 = .24$. A follow-up ANCOVA showed significant effects for retention, $F(4, 101) = 13.98$, $p < .001$, $\eta_p^2 = .36$, and transfer, $F(4, 101) = 9.80$, $p < .001$, $\eta_p^2 = .28$. Bonferroni-Holm-corrected pairwise comparisons for retention showed that the control group significantly performed worse than the group with positive and strongly connected pictures (difference: $M = 1.34$, $SE = 0.42$, $p = .016$, $\eta_p^2 = .19$) and better than the group with negative and weakly connected pictures (difference: $M = 1.58$, $SE = 0.45$, $p = .007$, $\eta_p^2 = .26$). Regarding transfer, comparisons showed that only the group with positive and strongly connected pictures performed significantly better than the control group (difference: $M = 2.00$, $SE = 0.49$, $p = .029$, $\eta_p^2 = .18$). In conclusion, pictures with a weakly connected and negative content can be seen as detrimental and pictures with a strongly connected and positive content as conducive (see also Figure 3) dependent from the learning scale.

Cognitive processes. Regarding further cognitive processes, an ANCOVA was conducted with the task-irrelevant thinking scores as dependent variables. The test revealed significant effects for affective charge, $F(1, 83) = 11.51$, $p = .001$, $\eta_p^2 = .12$, but not for connectedness, $F(1, 83) = 0.06$, $p > .05$, $\eta_p^2 < .01$, or the interaction, $F(1, 83) = 1.68$, $p > .05$, $\eta_p^2 = .02$. In conclusion, negative pictures increased task-irrelevant thoughts in contrast to positive pictures (see Table 1). An additional post hoc analysis of all five groups revealed that none of the picture groups differed significantly from the control group. In line with Question 3, affective charge was able to influence task-irrelevant thinking.

Discussion and Conclusion

The first experiment revealed that both a positive charge and a strong connection between text and decorative pictures increased

Table 1
Descriptive Results of Measures for Each Group in Experiment 1

	PS (n = 27)		PW (n = 17)		NS (n = 26)		NW (n = 19)		C (n = 19)	
Measure	M	SD	M	SD	M	SD	M	SD	M	SD
Text–picture relation	5.52	2.03	3.45	2.02	5.61	2.04	3.60	2.01		
Positive affect difference	.74	.99	.76	.99	−.24	.97	−.05	.96	.06	1.00
Negative affect difference	−.30	1.14	−.32	1.15	.79	1.12	.91	1.13	.49	1.13
Retention	**6.39**	1.40	5.58	1.40	4.60	1.38	**3.47**	1.39	5.05	1.39
Transfer	**5.87**	1.61	4.63	1.65	3.92	1.63	2.99	1.61	4.38	1.66
Task-irrelevant thinking	2.77	1.35	2.31	1.36	3.39	1.38	3.71	1.35	3.30	1.39

Note. Scores are adjusted for the following values of the covariates: prior knowledge = .33 and gender = 1.40. Mean scores of groups in bold text are significantly different than those of the control group. PS = positive and strongly connected pictures; PW = positive and weakly connected pictures; NS = negative and strongly connected pictures; NW = negative and weakly connected pictures; C = control group.

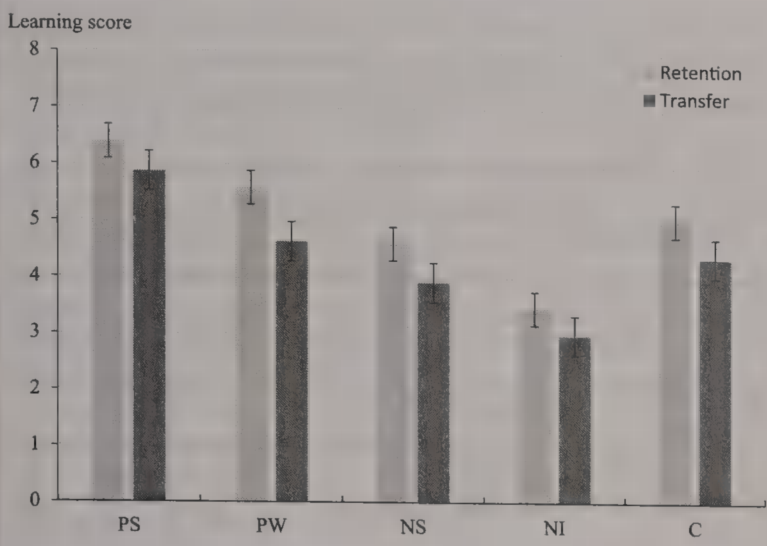

Figure 3. Retention and transfer scores of Experiment 1 by experimental groups. Retention and transfer scores ranged from zero to eight points. PS - positive and strongly connected pictures. PW - positive and weakly connected pictures. NS - negative and strongly connected pictures. NW - negative and weakly connected pictures. C - control group. Error bars indicate standard errors.

learning results with large effect sizes (Cohen, 1988). In addition, negative pictures induced a significantly higher assessment of task-irrelevant thinking than did positive pictures. These results are in line with those of previous experiments (Pekrun et al., 2002; Schneider et al., 2016). The pictures' connectedness to the text did not influence ratings of affective states so that both the flow hypothesis, which requires an increase in negative affect for weakly connected pictures (D'Mello & Graesser, 2012), and the fluency hypothesis, which requires an increase of positive affect for strongly connected pictures (Winkielman et al., 2003), appear not to be helpful in this context. Because connectedness also did not affect task-irrelevant thinking, a closer look at possible differences in cognitive facets (e.g., intrinsic and extraneous cognitive load; Kalyuga & Singh, 2016) would additionally help to evaluate the learning process. Although decorative pictures are supposed to increase extraneous processing and reduce learning, a positive valence and a stronger connectedness might decrease perceived difficulties (intrinsic cognitive load) and free resources for learning. Considering the results of the comparison with the control group, the use of positive and strongly connected decorative pictures seems reasonable but needs to be replicated under different conditions to ensure external validity. Thus, a second experiment was conducted to replicate findings with an alternative sample and to substantiate explanations for learning effects.

Experiment 2

Method

Participants and design. In total, 86 university students (74.4% female) from the Chemnitz University of Technology took part in this experiment. The mean age was 23.21 years (SD = 4.26). Students mainly majored in the field of media (66.3%) followed by psychology (14%), teaching (9.3%), and others (10.5%). Most of the students (91.9%) reported that their native language was German. All foreign students had a sufficient language level to fully understand learning material and questions asked in the experiment. Mean prior knowledge was 0.79 (SD = 1.26) out of seven points. Consistent with Experiment 1, students studied materials with either positive and strongly connected pictures (PS group; $N = 19$), negative and strongly connected pictures (NS group; $N = 17$), positive and weakly connected pictures (PW group; $N = 17$), or negative and weakly connected pictures groups (NW group; $N = 17$). Additionally, a fifth group, the control group (C) was given the same learning materials without decorative pictures ($N = 16$).

Materials and measures. The same learning web pages as described in Experiment 1 were used in this experiment. The same prior knowledge (α = .72), retention (α = .70) and transfer (α = .75), positive affect (α = .70) and negative affect (α = .76) questionnaires that were used in Experiment 1 were used in this experiment. As proposed in the discussion of Experiment 1, a closer look at cognitive processes would be helpful to explain how affective charge and connectedness result in a higher learning performance. For this, the cognitive load questionnaire by Leppink, Paas, van der Vleuten, van Gog, and van Merriënboer (2013) was included in addition to the test-irrelevant thinking scale. This questionnaire contained three sections, each rated on an 11-point scale ranging from 0 "totally incorrect" to 10 "totally correct"; however, our participants completed only the perceived intrinsic (ICL; α = .89) and extraneous (ECL; α = .79) cognitive load sections, as the germane cognitive load section was found to be theoretically (Kalyuga & Singh, 2016) and experimentally (Leppink et al., 2013) superfluous. Example items were·"The topics covered in the learning material were very complex" (ICL) and "The instructions and explanations within the learning material were very unclear" (ECL). These items were adapted to the text-based environment. A manipulation check item for connectedness and a demographic questionnaire were included as described in Experiment 1. In contrast, the demographic questionnaire did not include items asking for grade level or course profile because participants were students at a university. Instead, an item for major was included.

Procedure. Ten computers in a computer lab were prepared to display the starting page of the first questionnaire, and a participant number sheet was assigned to each workplace of the course room. The experimenter introduced all tasks and parts of the experiment with a premade instructions form to increase comparability. Students started with their experiments autonomously and continued as described in Experiment 1. All learners worked on the task for 5 min to 15 min. After the experiment, the participants had to sign a letter of agreement and were rewarded with either a study credit or €5 (5.64 Dollar).

Results

Data were analyzed, and results were similar to those of Experiment 1. No significant differences between the five groups of the experiment were found in terms of age, gender, grade, major, course credit, prior knowledge, time on task, perception of time adequacy, or on the positive activation and negative activation baseline scores ($ps > .05$). As a result, only prior knowledge was included as covariate, whereby only significant influences of the covariate were reported. All descriptive results are shown in Table 2.

Table 2
Descriptive Results of Measures for Each Group in Experiment 2

Measure	PS (n = 19) M	PS SD	PW (n = 17) M	PW SD	NS (n = 17) M	NS SD	NW (n = 17) M	NW SD	C (n = 16) M	C SD
Text–picture relation	5.63	1.74	4.13	1.73	5.54	1.73	2.70	2.19		
Positive affect difference	.51	1.13	.24	1.11	−.52	1.11	−.47	1.15	−.32	1.12
Negative affect difference	−.41	1.05	−.29	1.03	.79	1.03	.59	1.07	−.12	1.04
Retention	7.53	1.48	6.67	1.48	6.32	1.48	**5.26**	1.53	6.55	1.52
Transfer	**7.65**	1.39	6.41	1.40	6.17	1.40	5.95	1.40	6.47	1.40
Task-irrelevant thinking	3.02	1.13	3.11	1.15	3.96	1.15	3.95	1.15	3.27	1.16
Intrinsic cognitive load	2.86	1.53	3.60	1.57	3.59	1.57	**4.64**	1.57	3.09	1.56
Extraneous cognitive load	4.44	2.14	4.58	2.14	**6.55**	2.14	**6.45**	2.14	4.50	2.12

Note. Scores are adjusted for the following values of the covariates: prior knowledge = .79. Mean scores of groups in bold text are significantly different than those of the control group. PS = positive and strongly connected pictures; PW = positive and weakly connected pictures; NS = negative and strongly connected pictures; NW = negative and weakly connected pictures; C = control group.

Manipulation check. Again, a significant effect was found for connectedness, $F(1, 65) = 23.66$, $p < .001$, $\eta_p^2 = .29$, but not for affective charge, $F(1, 65) = 2.90$, $p > .05$, $\eta_p^2 = .05$, or the interaction, $F(1, 65) = 2.19$, $p > .05$, $\eta_p^2 = .04$.

Affective charge. Again, significant main effects in a MANCOVA were only found for affective charge, (Wilk's Λ = 0.75), $F(2, 64) = 10.70$, $p < .001$, $\eta_p^2 = .25$. Follow-up ANCOVAs for affective charge revealed significant effects for both the PAD, $F(1, 65) = 9.22$, $p = .003$, $\eta_p^2 = .12$, and the NAD scores, $F(1, 65) = 17.13$, $p < .001$, $\eta_p^2 = .21$. In line with Experiment 1, manipulation check was fully confirmed.

Learning performance. A MANCOVA for both learning scores revealed significant main effects for affective charge, (Wilk's Λ = 0.78), $F(2, 64) = 8.85$, $p < .001$, $\eta_p^2 = .22$, and for connectedness, (Wilk's Λ = 0.87), $F(2, 64) = 4.92$, $p = .010$, $\eta_p^2 = .13$, but not for the interaction, (Wilk's Λ = 0.96), $F(2, 64) = 1.41$, $p > .05$, $\eta_p^2 = .04$. Regarding retention, a follow-up ANCOVA revealed significant effects for both affective charge, $F(1, 65) = 12.37$, $p = .001$, $\eta_p^2 = .16$, and connectedness, $F(1, 83) = 6.60$, $p = .013$, $\eta_p^2 = .09$. Regarding transfer, significant effects were found for the affective charge, $F(1, 65) = 10.53$, $p = .002$, $\eta_p^2 = .14$, and connectedness, $F(1, 65) = 6.14$, $p = .016$, $\eta_p^2 = .09$. In sum, positive or strongly connected decorative pictures fostered retention and transfer performance in contrast to negative and weakly connected pictures (see Table 2). These results are in line with Experiment 1.

Differences to the control group. A subsequent MANCOVA with all five groups revealed a significant main effect, (Wilk's Λ = 0.72), $F(8, 158) = 3.48$, $p = .001$, $\eta_p^2 = .15$, and a significant effect of prior knowledge, (Wilk's Λ = 0.85), $F(2, 79) = 6.92$, $p = .002$, $\eta_p^2 = .15$. Follow-up ANCOVAs show significant effects for retention, $F(4, 80) = 5.19$, $p = .001$, $\eta_p^2 = .21$, and transfer, $F(4, 80) = 4.07$, $p = .005$, $\eta_p^2 = .17$. Bonferroni-Holm-corrected pairwise comparisons for retention showed that the control group only performed significantly better than the group with negative and weakly connected pictures (difference: $M = 1.29$, $SE = 0.52$, $p = .011$, $\eta_p^2 = .16$). Regarding transfer, comparisons showed that only the group with positive and strongly connected pictures performed significantly better than the control group (difference: $M = 1.19$, $SE = 0.48$, $p = .010$, $\eta_p^2 = .16$). In conclusion, pictures with a weak connectedness and a negative content can be seen as detrimental and pictures with a strong connectedness and positive content as conducive depending on the learning scores (see also Figure 4).

Cognitive processes. Regarding further cognitive processes, an ANCOVA was conducted for task-irrelevant thinking. The test revealed significant effects for affective charge, $F(1, 65) = 10.57$, $p = .002$, $\eta_p^2 = .14$, but not for connectedness, $F(1, 65) = 0.02$, $p > .05$, $\eta_p^2 < .01$, or the interaction, $F(1, 65) = 0.05$, $p > .05$, $\eta_p^2 = .02$. In conclusion, negative pictures increased task-irrelevant thoughts in contrast to positive pictures (see Table 2).

A MANCOVA for both cognitive load scores revealed significant main effects for affective charge, (Wilk's Λ = 0.71), $F(2, 64) = 13.21$, $p < .001$, $\eta_p^2 = .29$, and for connectedness, (Wilk's Λ = 0.91), $F(2, 64) = 3.02$, $p = .048$, $\eta_p^2 = .09$, but not for the interaction, (Wilk's Λ = 0.99), $F(2, 64) = 0.12$, $p > .05$, $\eta_p^2 < .01$.

Regarding ICL, a follow-up ANCOVA revealed a significant effect for affective charge, $F(1, 65) = 5.73$, $p = .020$, $\eta_p^2 = .08$, and connectedness, $F(1, 65) = 5.86$, $p = .018$, $\eta_p^2 = .08$. Regarding ECL, a significant effect was only found for affective charge, $F(1, 65) = 16.25$, $p < .001$, $\eta_p^2 = .20$. In sum, positive or weakly

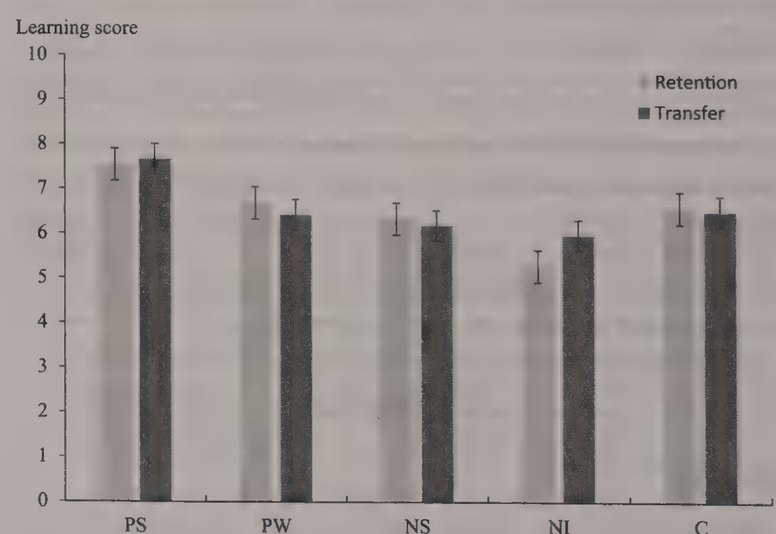

Figure 4. Retention and transfer scores by experimental groups of Experiment 2. Retention and transfer scores ranged from zero to eight points. PS - positive and strongly connected pictures. PW - positive and weakly connected pictures. NS - negative and strongly connected pictures. NW - negative and weakly connected pictures. C - control group. Error bars indicate standard errors.

connected decorative pictures decrease a perception of ICL compared with negative or weakly connected pictures. In addition, ECL was lowered for positive pictures in contrast to negative pictures (see Table 2).

An additional post hoc analysis of all five groups concerning task-irrelevant thinking revealed that none of the picture groups differ significantly from the control group. A post hoc analysis of all five groups regarding both cognitive load scores revealed that negative and weakly connected pictures were assessed significantly higher in ICL than the control group (difference: $M = 1.55$, $SE = 0.54$, $p = .005$, $\eta_p^2 = .21$). In addition, negative strongly connected pictures (difference: $M = 2.04$, $SE = 0.74$, $p = .007$, $\eta_p^2 = .20$) and negative weakly connected pictures (difference: $M = 1.95$, $SE = 0.74$, $p = .010$, $\eta_p^2 = .18$) were rated significantly higher in ECL than in the control group.

Discussion and Conclusion

The second experiment showed similar results in comparison to the first experiment. Again, positive and strongly connected pictures supported learning and negative pictures, in particular increased task-irrelevant thoughts. In addition to the first experiment, the analysis of cognitive load facets revealed that positive pictures also increased the assessment of ECL whereas ICL scores were significantly decreased. Strongly connected pictures were also found to decrease the perception of ICL. As a consequence, negative and weakly connected pictures were rated significantly higher in ICL compared with the control group and all groups with negatively charged pictures scored significantly higher in their ECL than the control group. These results are in line with previous studies (e.g., Pekrun et al., 2002; Schneider et al., 2016) and support the advantageous imagination hypothesis (Leahy & Sweller, 2008) or the high cohesion hypothesis (Sukalla et al., 2015). In line with the first experiment, the same configurations of decorative pictures were detected as conducive or detrimental, although both learning scores were affected differently. This might be a result of the new sample of Experiment 2. In addition, results can only be generalized to similar learning topics. As a consequence, the third experiment was conducted to be able to make findings more generalizable (new learning topic).

Experiment 3: Prestudy

Method

Participants and design. The prestudy was conducted similarly to the prestudy of Experiment 1. In total, 42 participants (61.9% female) voluntarily took part in the prestudy. The mean age was 25.6 years ($SD = 4.3$). Most of the students (85.7%) reported their native language to be German, and all foreign students had a sufficient language level.

Materials, measures, and procedure. To keep the structure consistent with the first two experiments, pictures from three categories (human brain, nutrition, and movement) were selected from web resources. To create differences in the affective charge of decorative pictures, anthropomorphic features (e.g., mouth and eyes) were added to topic-related objects, resulting in either positive or negative faces. For instance, an angry face was added to an illustration of an orange to elicit states of negative affect (e.g., frustration and anger). In total, 53 pictures were selected, whereby 14 showed anthropomorphized brains (six to evoke positive affect, eight to evoke negative affect), 13 depicted anthropomorphized food (six to evoke positive affect, seven to evoke negative affect), and 14 showed athletes (seven for each affective dimension). Consistent with the prestudy of Experiment 1, the same emotional data (positive and negative affect) and demographical data (age, sex, and native language) were collected.

Results

The results were analyzed consistently to the prestudy of Experiment 1. Regarding positive affect, results showed a significant main effect, (Wilk's $\Lambda = 0.76$), $F(3, 12) = 43.99$, $p < .001$, $\eta_p^2 = .76$. Positive pictures ($M = 4.51$, $SD = 0.99$) were assessed as more positive than negative pictures ($M = 3.13$, $SD = 0.51$). The interaction terms of positive affect and all included between-subjects factors were not significant ($p = .14, .90$). Follow-up tests showed significant differences in terms of positive affect for all three distinct categories of pictures, namely brain, $F(1, 9) = 13.75$, $p = .002$, $\eta_p^2 = .50$, nutrition, $F(1, 10) = 8.71$, $p = .011$, $\eta_p^2 = .38$, and movement, $F(1, 23) = 14.01$, $p = .001$, $\eta_p^2 = .50$.

Regarding negative affect, results revealed a significant main effect, (Wilk's $\Lambda = 0.2$), $F(3, 12) = 56.61$, $p < .001$, $\eta_p^2 = .80$. Negative pictures ($M = 4.76$, $SD = 1.18$) were assessed as more negative than positive pictures ($M = 3.14$, $SD = 0.52$). The interaction of negative affect and all included between-subjects factors was not significant, $p = [.07, .78]$. Follow-up tests showed significant differences in terms of negative affect for all three distinct categories of pictures, namely brain, $F(1, 21) = 50.01$, $p < .001$, $\eta_p^2 = .78$, nutrition, $F(1, 18) = 28.34$, $p < .001$, $\eta_p^2 = .67$, and movement, $F(1, 28) = 55.34$, $p < .001$, $\eta_p^2 = .80$.

These results revealed that the preselected positive and negative pictures evoked the respective affective charge. Consequently, among the pictures with the highest scores in terms of corresponding affective dimension per category, two illustrations with different motifs were chosen (see Figure 5).

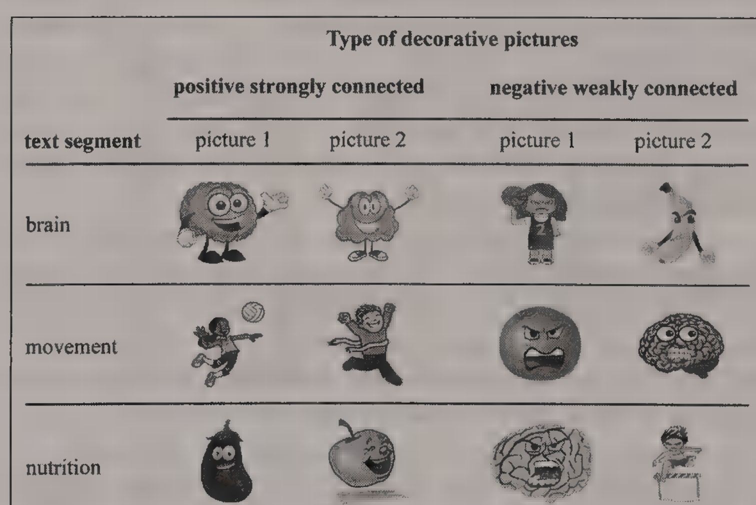

Figure 5. Experimental pictures used in Experiment 3 together with a learning text about the human body. See the online article for the color version of this figure.

Experiment 3

Method

Participants and design. Overall, 162 secondary school students (58.6% female) from the same school as in Experiment 1 took part in this experiment. The mean age was 17.56 years (SD = 1.14). Students were either in Grade 11 (40.1%), 12 (37.7%), or 13 (22.2%) and majored in either economy (39.5%), design and media (35.8%), or health and social issues (24.7%). One student (0.6%) reported that his native language was not German, but he showed a high German language proficiency. Mean prior knowledge was 0.66 out of 3 points (SD = 0.35). Consistent with Experiments 1 and 2, students studied materials with either positive and strongly connected pictures (PS group; N = 33), negative and strongly connected pictures (NS group; N = 32), positive and weakly connected pictures (PW group; N = 33), or negative and weakly connected pictures groups (NW group; N = 32). Additionally, a fifth group, the control group (C) was given the same learning materials without decorative pictures (N = 32).

Materials and measures. The learning material consisted of 1,518 words and either zero (C) or six decorative pictures, which were chosen based on the prestudy, depending on the experimental condition. The learning text consists of natural scientific facts about the human body and was separated into three segments: "brain," "nutrition," and "movement." In the "brain" section, students learned biological facts about the structure and different parts of the brain and its neurons. In the "food" section, nutritional scientific facts about different kinds of nutrients and the energy balance were provided. In the "movement" section, learners studied facts about the human muscle system, its generation of energy, and cardio training.

Each segment was separated into two chapters as described in Experiment 1. Except for the mentioned differences, the learning environment was consistent with that of Experiments 1 and 2.

Measures. A prior knowledge questionnaire was created. This questionnaire consisted of three questions in an open-answer format (one for each category), which aimed at measuring the level of knowledge concerning the topic. The questions were (a) "What are the main parts of the brain?", (b) "What is the sports scientific term for a muscle working without oxygen?", and (c) "Name all nutrients!" The first task was rewarded with 0.25 points per correct answer resulting in a maximum of one point. Question 2 was rewarded with one point if the correct answer was given. Question 3 was rewarded with 0.16 points per correct answer resulting in a maximum of one points. In total, three points could be reached with all three prior-knowledge questions (α = .67).

A second task was designed to measure retention knowledge. This questionnaire consisted of 14 single-choice questions, comparable to those used in Experiments 1 and 2. For example, students were given the answer possibilities "6 cm", "12 cm", "18 cm", and "24 cm" to answer the question, "What is the maximum length of a muscle fiber?" If the correct answer was given, learners were rewarded with one point per correct answer. A maximum of 14 points could be reached regarding this knowledge category (α = .72).

A third task was created to measure transfer knowledge. This questionnaire consisted of 11 single-choice questions. All transfer questions aimed at applying recently achieved knowledge about the human body within new situations. One example question in this category was "Regarding the shape of the following examples, which is most likely shaped like a neuron?" Participants were given the following answer possibilities: "a multistory building," "a tree," "a coffee mug," and "a baseball cap." For this question, students had to apply their previously gained knowledge about the shape of neurons with axons and dendrites. Again, each question was rewarded with one point. A maximum of 11 points could be reached for all transfer questions (α = .74).

Consistent with Experiments 1 and 2, the emotional dimensions of learners' positive affect (α = .70) and negative affect (α = .76) were measured by using the PANAVA-KS questionnaire (Schallberger, 2005), and, similar to Experiment 2, learners' cognitive load was measured with the cognitive load questionnaire by Leppink and colleagues (2013). A manipulation check item for connectedness, two items measuring perceived time adequacy, and a demographic questionnaire were included as described in Experiment 1.

Procedure. To maintain high comparability between the experiments, the procedure was kept similar to that as described in Experiment 1. Thus, the experiment took place in the same classroom as in Experiment 1. Nevertheless, one change had to be made: Because of an Internet connectivity problem, participants completed a paper version of the questionnaire and viewed the learning pages for each experimental condition (i.e., PC, NC, PI, NI, and C) via randomly distributed CDs. The starting page of the first questionnaire was launched on each computer at the beginning of each lesson. Students first completed questionnaires collecting data on their prior knowledge and current emotional states. At the end of this part, participants were instructed to wait until the experimenter explained how to continue. When all students finished the first part, they were told to start with the learning pages. The students continued autonomously to the third part, which included all dependent variables or covariates measured via a second paper questionnaire. On average, participants took 10.19 min (SD = 2.93) to complete the learning pages. The experiment was conducted in 4 school days. Class sizes differed between 13 and 23 students (M = 20.25, SD = 3.65).

Results

Data were analyzed, and results were similar to those in Experiments 1 and 2. There were no significant differences between the five experimental groups in terms of age, gender, grade, course of study, starting time, day of experiment, perception of time adequacy, and the baseline measurements of positive and negative affect ($ps > .05$). As a result, only prior knowledge was included as covariate, whereby only significant influences of the covariate were reported. All descriptive results are shown in Table 3.

Manipulation check. Again, a significant effect was found for connectedness, $F(1, 125) = 42.92$, $p < .001$, $\eta_p^2 = .25$, but not for affective charge, $F(1, 125) = 3.01$, $p > .05$, $\eta_p^2 = .02$, or the interaction, $F(1, 125) < 0.01$, $p > .05$, $\eta_p^2 < .01$.

Affective charge. Again, significant main effects in a MANCOVA were found only for affective charge, (Wilk's Λ = 0.76), $F(2, 124) = 14.23$, $p < .001$, $\eta_p^2 = .25$. Follow-up ANCOVAs for affective charge revealed significant effects for both the PAD, $F(1, 125) = 31.15$, $p < .001$, $\eta_p^2 = .20$, and the NAD scores, $F(1, 125) = 28.19$, $p < .001$, $\eta_p^2 = .18$. Supporting Experiments 1 and

Table 3
Descriptive Results of Measures for Each Group in Experiment 3

Measure	PS (n = 33)		PW (n = 33)		NS (n = 32)		NW (n = 32)		C (n = 32)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Text–picture relation	5.10	1.78	3.08	1.78	4.57	1.75	2.52	1.75	—	—
Positive affect difference	.63	1.61	1.07	1.61	−.97	1.64	−.74	1.58	−.54	1.58
Negative affect difference	−.29	1.49	−.87	1.49	.97	1.47	.87	1.47	.17	1.47
Retention	**8.69**	2.36	7.67	2.36	6.43	2.38	**5.24**	2.38	6.86	2.38
Transfer	**7.69**	1.95	**6.06**	1.95	4.76	1.92	4.89	1.92	4.65	1.92
Intrinsic cognitive load	**3.03**	2.47	4.16	2.64	4.36	2.49	5.80	2.49	5.30	2.49
Extraneous cognitive load	**4.94**	1.61	4.38	1.67	**5.54**	1.58	**6.12**	1.58	3.75	1.58

Note. Scores are adjusted for the following values of the covariates: prior knowledge = .66. Mean scores of groups in bold text are significantly different than those of the control group. PS = positive and strongly connected pictures; PW = positive and weakly connected pictures; NS = negative and strongly connected pictures; NW = negative and weakly connected pictures; C = control group.

2, students with strongly connected pictures reported higher scores of connectedness than did students with weakly connected pictures. In addition, students with negative pictures reported lower scores of PAD and higher scores of NAD than did students with positive pictures (see Table 3).

Learning performance. A MANCOVA for both factors revealed significant main effects for affective charge, (Wilk's Λ = 0.72), $F(2, 124) = 23.60$, $p < .001$, $\eta_p^2 = .28$, and for connectedness, (Wilk's $\Lambda = 0.94$), $F(2, 124) = 4.25$, $p = .016$, $\eta_p^2 = .06$, and for the interaction, (Wilk's $\Lambda = 0.94$), $F(2, 124) = 4.32$, $p = .015$, $\eta_p^2 = .07$. Regarding retention, a follow-up ANCOVA revealed significant effects for affective charge, $F(1, 125) = 30.51$, $p < .001$, $\eta_p^2 = .20$, and connectedness, $F(1, 125) = 6.93$, $p = .010$, $\eta_p^2 = .05$, and for the interaction, $F(1, 125) = 0.04$, $p > .05$, $\eta_p^2 < .01$. Regarding transfer, significant effects were found for affective charge, $F(1, 125) = 35.86$, $p < .001$, $\eta_p^2 = .22$, and for the interaction, $F(1, 125) = 6.84$, $p = .010$, $\eta_p^2 = .05$, but not for connectedness, $F(1, 125) = 4.94$, $p = .028$, $\eta_p^2 = .04$. In sum, positive or strongly connected decorative pictures fostered retention and transfer performance in contrast to negative or weakly connected pictures. Regarding the interaction, positive decorative pictures were shown to be not significantly beneficial in contrast to negative pictures unless they were weakly connected (see Table 3). Except for the interaction effect, the results are in line with those of Experiments 1 and 2.

Differences to the control group. A subsequent MANCOVA with all five groups revealed a significant main effect, (Wilk's Λ = 0.65), $F(8, 310) = 9.32$, $p < .001$, $\eta_p^2 = .19$. Follow-up ANCOVAs showed significant effects for retention, $F(4, 156) = 9.81$, $p < .001$, $\eta_p^2 = .20$, and transfer, $F(4, 156) = 14.68$, $p < .001$, $\eta_p^2 = .27$. Bonferroni-Holm-corrected pairwise comparisons for retention revealed that the control group significantly performed worse than the group with positive and strongly connected pictures (difference: $M = 1.82$, $SE = 0.58$, $p = .002$, $\eta_p^2 = .13$) but better than the group with negative and weakly connected pictures (mean difference = 1.62, $SE = 0.59$, $p = .007$, $\eta_p^2 = .11$). Regarding transfer, comparisons showed that only the groups with positive and strongly connected pictures (difference: $M = 3.04$, $SE = 0.48$, $p < .001$, $\eta_p^2 = .39$) and with positive and weakly connected pictures (difference: $M = 1.41$, $SE = 0.48$, $p = .004$, $\eta_p^2 = .12$) performed significantly better than the control group. In conclusion, pictures with a weak connectedness and negative content can be seen as detrimental and pictures with a strong connectedness and positive content as conducive (see also Figure 6).

Cognitive processes. A MANCOVA for both cognitive load scores revealed significant main effects for affective charge, (Wilk's $\Lambda = 0.86$), $F(2, 121) = 8.49$, $p < .001$, $\eta_p^2 = .15$, and for connectedness, (Wilk's $\Lambda = 0.94$), $F(2, 121) = 3.54$, $p = .020$, $\eta_p^2 = .06$, but not for the interaction, (Wilk's $\Lambda = 0.97$), $F(2, 121) = 2.48$, $p > .05$, $\eta_p^2 = .03$.

Regarding ICL, a follow-up ANCOVA revealed significant effects for both affective charge, $F(1, 122) = 9.62$, $p = .002$, $\eta_p^2 = .07$, and connectedness, $F(1, 122) = 7.48$, $p = .007$, $\eta_p^2 = .06$. Regarding ECL, a significant effect was found only for the affective charge of decorative pictures, $F(1, 122) = 16.74$, $p < .001$, $\eta_p^2 = .12$. In sum, positive decorative pictures decreased the perception of ICL and ECL in contrast to negative pictures. In addition, strongly connected pictures decreased the perception of ICL in contrast to weakly connected pictures (see Table 3).

A post hoc analysis of all five groups regarding all cognitive load scores revealed that positive and strongly connected pictures

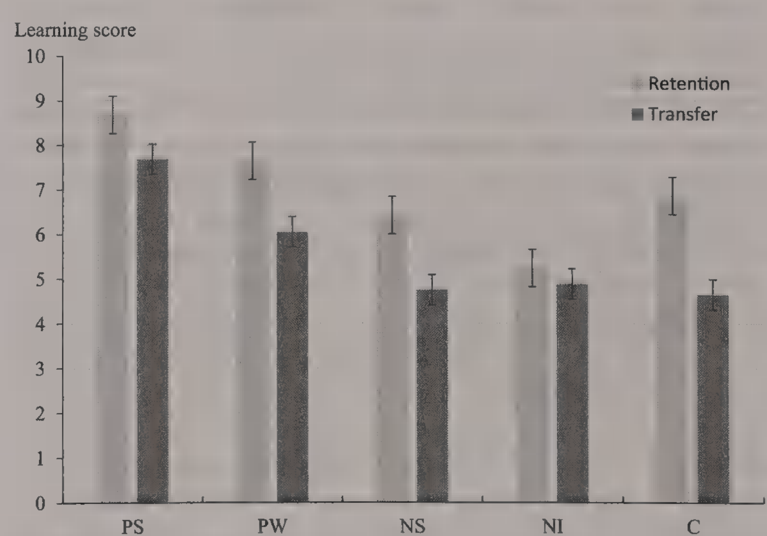

Figure 6. Retention and transfer scores by experimental groups of Experiment 3. Retention scores ranged from zero to 14 points and transfer scores ranged from zero to eleven points. PS - positive and strongly connected pictures. PW - positive and weakly connected pictures. NS - negative and strongly connected pictures. NW - negative and weakly connected pictures. C - control group. Error bars indicate standard errors.

were assessed significantly lower in ICL than in the control group (difference: $M = 2.27$, $SE = 0.61$, $p < .001$, $\eta_p^2 = .18$). In addition, negative weakly connected (difference: $M = 2.37$, $SE = 0.40$, $p < .001$, $\eta_p^2 = .37$), negative strongly connected (difference: $M = 1.79$, $SE = 0.40$, $p < .001$, $\eta_p^2 = .25$), and positive strongly connected pictures (difference: $M = 1.19$, $SE = 0.39$, $p = .003$, $\eta_p^2 = .13$) were rated significantly higher in ECL than the control group.

Discussion and Conclusion

The aim of the third experiment was to investigate whether the results of Experiments 1 and 2 can be generalized independent from the learning topic. Again, learning results increased when decorative pictures were either positive or strongly connected to a learning text rather than negative or weakly connected. In contrast to the results of Experiments 1 and 2, an interaction for transfer was shown revealing that positive pictures were more effective when they were also strongly connected to the text. Such a pattern was also observed in Experiments 1 and 2 on a descriptive level but did not reach statistical significance, which is probably because of power issues. A second contrast was shown for the examination of differences between the group without decorative pictures and all other experimental conditions. In this case, both positive and strongly connected pictures and positive and weakly connected pictures were found to be conducive. Again, the descriptive results of Experiments 1 and 2 showed similar patterns, whereby significance was not reached. In line with Experiments 1 and 2, a strong connection of decorative pictures reduced the perception of ICL, whereas positive pictures reduced the assessment of ECL. In conclusion, the results of Experiment 3 underline the importance of affective charge as well as connectedness as boundary conditions of the effectiveness of decorative pictures in multimedia learning.

General Discussion

In the present study, we aimed to examine the role of decorative pictures' affective charge and connectedness as two potential boundary conditions for learning. In line with a previous study (Schneider et al., 2016), positive decorative pictures fostered both retention and transfer performance in contrast to pictures with a negative charge. The higher reported test-irrelevant thinking (measured in Experiments 1 and 2) and extraneous cognitive load scores (according to Experiments 2 and 3) for the negative pictures condition might contribute to the detrimental effect of negative emotions. Besides emotions, a strong connectedness of decorative pictures to the text aided learning in all three experiments, whereas a weak semantic relation decreased learning outcomes. Both a positive charge and a strong text–picture interconnectedness resulted in lower perceived intrinsic load scores compared with the contrary conditions, leading to the assumption that both moderators might reduce the assessed task complexity. Because learners' emotional states were consistently influenced by the affective charge of pictures but not by their connectedness, both the presumed flow and fluency hypothesis might not be applicable the context of decorative pictures. In contrast, the further postulated assumptions (on high cohesion and advantageous imagination) remain as possible explanations. In addition, an interaction effect of both features was encountered in only one case. In Experiment 3, learners' transfer performance was increased when provided with positive and strongly connected pictures, leading to the assumption that these pictures foster learning beyond a pure recall of information. Nonsignificant findings regarding this interaction in Experiments 1 and 2 might be explained by insufficient statistical power or, perhaps, by different degrees of the pictures' groundedness (Belenky & Schalk, 2014). That is, the transfer of knowledge to novel problems might be particularly fostered when the design of the decorative pictures is more abstract because subsequent mental representation will be more general, less related to a specific situation, and therefore easier to apply.

Nevertheless, the comparison of the four decorative picture conditions with the text-only group emphasizes the interplay of both examined features, showing that the control group was outperformed by the positive and strongly connected pictures condition (only in Experiments 1 and 3) but achieved higher scores than the negative and weakly connected pictures condition in terms of retention performance. Regarding transfer, both positive pictures groups achieved higher scores than the control group, although the comparison with the weakly interconnected pictures condition only reached statistical significance in the third experiment. Together with the higher extraneous cognitive load reported by learners who were provided with negative pictures (compared with those in the text-only condition), these findings indicate that negative and weakly connected decorative pictures might be considered extraneous materials. According to the coherence principle (Mayer, 2014), such pictures should be excluded.

Implications

The present study contributes to the ongoing discussion on the role of decorative pictures in multimedia message processing. To integrate the contradicting views that result from inconsistent findings in previous research, we propose a framework of boundary conditions that presumably moderate such pictures' impact on learning outcomes. The results of three experiments provide further empirical evidence regarding the assumption that decorative pictures' effects on learning outcomes should not be considered detrimental in general but are likely to be determined by their specific design (see also Danielson et al., 2015; Schneider et al., 2016). Our findings suggest that interspersed decorative pictures are conducive to the learning process when such illustrations elicit positive emotional states and foster cognitive processing by a strong connectedness to the presented text. In contrast, when decorative pictures evoke negative affect or are only weakly connected to the text, extraneous cognitive load is imposed by task-irrelevant thoughts or superfluous integrative attempts. In line with the coherence principle (Mayer, 2014), such pictures can be considered extraneous details that hinder learning.

On the basis of these results, we recommend that instructional material designers take both affective charge and connectedness into account when including decorative depictions or illustrations in textbooks or multimedia learning materials. In contrast to previous assumptions (e.g., Harp & Mayer, 1998), we urge designers to add not only instructive but also positive and text-related decorative pictures to learning materials because the benefit yielded by such elements is likely to surpass the costs caused by having to process additional elements. The inclusion of such aesthetically

appealing representations appears appropriate because human learning and performance cannot be described from a mere cognitive perspective that neglects affective responses to the information we perceive (Plass & Kaplan, 2015).

Limitations and Future Directions

In the present study, highly reliable, subject-reliant questionnaires were applied to measure learners' affect, cognition, and motivation. Future studies should additionally include psychophysiological measurements, such as electroencephalography (Shen, Wang, & Shen, 2009), to substantiate findings. Knowledge retention and transfer were measured immediately after the affective and cognitive questionnaires. Furthermore, this nonrandomized presentation of numerous questionnaires may have resulted in sequencing biases. In future studies, a delayed posttest would reveal possible long-term effects (Schweppe & Rummer, 2014). Because the impact of decorative pictures on learning might also vary in terms of their degree of interestingness (Mayer et al., 2008), further studies should investigate whether positive and negative affective pictures differ in this regard. In addition, recent research indicates that irrelevant pictures in learning materials are ignored with increasing experience (Rop, van Wermeskerken, de Nooijer, Verkoeijen, & van Gog, 2016). Consequently, further studies should include eye-tracking methods to investigate whether learners' eye movements differ according to the investigated boundary conditions as a function of task experiences. A possible interpretation of differences in the groundedness of decorative pictures might also be analyzed by systematic variation. To examine whether descriptive differences might become significant, future studies should increase the sample size to be able to reveal small to medium effects.

From a theoretical perspective, we proposed various hypotheses regarding the impact of decorative pictures' degree of connectedness (e.g., the high cohesion hypothesis). Revealed impacts on learning outcomes might serve as a first indicator for their validity. However, more detailed measures (e.g., eye-tracking) should be conducted in order to provide deeper insights into learners' relational reasoning (Danielson & Sinatra, 2016). Besides cognition, educational researchers stress that emotions are inherently motivational, that is that emotions might exert powerful motivational influences (e.g., Pekrun, 2006; Plass & Kaplan, 2015). Consequently, future research might benefit from taking motivational influences into account, for example measurements of learners' initial motivation as well as motivational changes caused by the pictures' design.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13,* 219–235. http://dx.doi.org/10.1177/1088868309341564

Belenky, D. M., & Schalk, L. (2014). The effects of idealized and grounded materials on learning, transfer, and interest: An organizing framework for categorizing external knowledge representations. *Educational Psychology Review, 26,* 27–50. http://dx.doi.org/10.1007/s10648-014-9251-9

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1,* 276–298. http://dx.doi.org/10.1037/1528-3542.1.3.276

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review, 14,* 5–26. http://dx.doi.org/10.1023/A:1013176309260

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *The British Journal of Educational Psychology, 62,* 233–246. http://dx.doi.org/10.1111/j.2044-8279.1992.tb01017.x

Chang, Y., & Choi, S. (2014). Effects of seductive details evidenced by gaze duration. *Neurobiology of Learning and Memory, 109,* 131–138. http://dx.doi.org/10.1016/j.nlm.2014.01.005

Chen, X., & Latham, G. P. (2014). The effect of priming learning vs. performance goals on a complex task. *Organizational Behavior and Human Decision Processes, 125,* 88–97. http://dx.doi.org/10.1016/j.obhdp.2014.06.004

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Science* (2nd ed.). Hillsdale, NJ: Erlbaum.

Danielson, R. W., Schwartz, N. H., & Lippmann, M. (2015). Metaphorical graphics aid learning and memory. *Learning and Instruction, 39,* 194–205. http://dx.doi.org/10.1016/j.learninstruc.2015.07.004

Danielson, R. W., & Sinatra, G. M. (2016). A relational reasoning approach to text-graphic processing. *Educational Psychology Review, 29,* 1–18.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22,* 145–157. http://dx.doi.org/10.1016/j.learninstruc.2011.10.001

Gläser-Zikuda, M., Fuß, S., Laukenmann, M., Metz, K., & Randler, C. (2005). Promoting students' emotions and achievement—Instructional design and evaluation of the ECOLE-approach. *Learning and Instruction, 15,* 481–495. http://dx.doi.org/10.1016/j.learninstruc.2005.07.013

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90,* 414–434. http://dx.doi.org/10.1037/0022-0663.90.3.414

Heidig, S., Müller, J., & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior, 44,* 81–95. http://dx.doi.org/10.1016/j.chb.2014.11.009

Huk, T., & Ludwigs, S. (2009). Combining cognitive and affective support in order to promote learning. *Learning and Instruction, 19,* 495–505. http://dx.doi.org/10.1016/j.learninstruc.2008.09.001

Isen, A. M., Daubman, K. A., & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology, 52,* 1122–1131. http://dx.doi.org/10.1037/0022-3514.52.6.1122

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 576–597). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.028

Kalyuga, S., & Singh, A. M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review, 28,* 831–852. http://dx.doi.org/10.1007/s10648-015-9352-0

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin, 139,* 917–940. http://dx.doi.org/10.1037/a0030811

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174. http://dx.doi.org/10.2307/2529310

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instructional manual.* [Tech. Rep. No. A-8]. Gainesville, FL: University of Florida, National Institute of Mental Health Center for the Study of Emotion and Attention.

Leahy, W., & Sweller, J. (2008). The imagination effect increases with an increased intrinsic cognitive load. *Applied Cognitive Psychology, 22,* 273–283. http://dx.doi.org/10.1002/acp.1373

Lee, Y., & Nelson, D. (2004). Instructional use of visual representations of knowledge. In R. Ferdig, C. Crawford, R. Carlsen, N. Davis, J. Price, R. Weber, & D. Willis (Eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference* (pp. 2371–2378). Chesapeake, VA: Association for the Advancement of Computing in Education.

Lenzner, A., Schnotz, W., & Müller, A. (2013). The role of decorative pictures in learning. *Instructional Science, 41,* 811–831. http://dx.doi.org/10.1007/s11251-012-9256-z

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45,* 1058–1072. http://dx.doi.org/10.3758/s13428-013-0334-1

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. In B. K. Britton, A. Woodward, & M. Binkley (Eds.), *Learning from textbooks: Theory and practice* (pp. 95–119). Hillsdale, NJ: Lawrence Erlbaum.

Levin, J. R. (1981). On the functions of pictures in prose. In F. J. Pirozzolo & M. C. Wittrock (Eds.), *Neuropsychological and cognitive processes in reading* (pp. 203–228). San Diego, CA: Academic Press.

Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A grumpy overview. *Social Cognition, 22,* 530–554. http://dx.doi.org/10.1521/soco.22.5.530.50767

Linderholm, T., Everson, M. G., Van Den Broek, P., Mischinski, M., Crittenden, A., & Samuels, J. (2000). Effects of causal text revisions on more-and less-skilled readers' comprehension of easy and difficult texts. *Cognition and Instruction, 18,* 525–556. http://dx.doi.org/10.1207/S1532690XCI1804_4

Magner, U. I., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction, 29,* 141–152. http://dx.doi.org/10.1016/j.learninstruc.2012.07.002

Marsh, E. E., & White, M. D. (2003). A taxonomy of relationships between images and texts. *The Journal of Documentation, 59,* 647–672. http://dx.doi.org/10.1108/00220410310506303

Mayer, R. E. (1993). Comprehension of graphics in texts: An overview. *Learning and Instruction, 3,* 239–245. http://dx.doi.org/10.1016/0959-4752(93)90007-M

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.005

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.005

Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied, 14,* 329–339. http://dx.doi.org/10.1037/a0013835

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19,* 309–326. http://dx.doi.org/10.1007/s10648-007-9047-2

Nadler, R. T., Rabi, R., & Minda, J. P. (2010). Better mood and better performance. Learning rule-described categories is enhanced by positive mood. *Psychological Science, 21,* 1770–1776. http://dx.doi.org/10.1177/0956797610387441

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior, 44,* 267–278. http://dx.doi.org/10.1016/j.chb.2014.10.061

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education, 86,* 30–42. http://dx.doi.org/10.1016/j.compedu.2015.02.016

Park, B., Moreno, R., Seufert, T., & Brünken, R. (2011). Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behavior, 27,* 5–10. http://dx.doi.org/10.1016/j.chb.2010.05.006

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18,* 315–341. http://dx.doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37,* 91–105. http://dx.doi.org/10.1207/S15326985EP3702_4

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction, 29,* 128–140. http://dx.doi.org/10.1016/j.learninstruc.2013.02.006

Plass, J. L., & Kaplan, U. (2015). Emotional design in digital media for learning. In S. Tettegah & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–162). New York, NY: Elsevier.

Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology, 17,* 715–734. http://dx.doi.org/10.1017/S0954579405050340

Pozzer, L. L., & Roth, W.-M. (2003). Prevalence, function and structure of photographs in high school biology textbooks. *Journal of Research in Science Teaching, 40,* 1089–1114. http://dx.doi.org/10.1002/tea.10122

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7,* 216–237. http://dx.doi.org/10.1016/j.edurev.2012.05.003

Rey, G. D. (2014). Seductive details and attention distraction–An eye tracker experiment. *Computers in Human Behavior, 32,* 133–144. http://dx.doi.org/10.1016/j.chb.2013.11.017

Rop, G., Van Wermeskerken, M., De Nooijer, J. A., Verkoeijen, P. P. J. L., & Van Gog, T. (2016). Task experience as a boundary condition to the negative effects of irrelevant information on learning. *Educational Psychology Review*. Advance online publication. http://dx.doi.org/10.1007/s10648-016-9388-9

Roseman, I. J. (2011). Emotional behaviors, emotivational goals, emotion strategies: Multiple levels of organization integrate variable and consistent responses. *Emotion Review, 3,* 434–443. http://dx.doi.org/10.1177/1754073911410744

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172. http://dx.doi.org/10.1037/0033-295X.110.1.145

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition, 34,* 344–355. http://dx.doi.org/10.3758/BF03193412

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46,* 929–938. http://dx.doi.org/10.1037/0022-3514.46.4.929

Schallberger, U. (2005). *Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling-Studien (PANAVA-KS). Forschungsberichte aus dem Projekt: "Qualität des Erlebens in Arbeit und Freizeit", 6* [Short sclaes to measure the positive activation, negative activation, and valence in experience sampling studies (PANAVA-SS). Research report from the project: "Quality of experiences in work and leisure time."]. Zürich, Switzerland: Psycholo-

gisches Institut der Universität Zürich, Fachrichtung Angewandte Psychologie.

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Sciences Information, 44,* 695–729. http://dx.doi.org/10.1177/0539018405058216

Schneider, S., Nebel, S., & Rey, G. D. (2016). Decorative pictures and emotional design in multimedia learning. *Learning and Instruction, 44,* 65–73. http://dx.doi.org/10.1016/j.learninstruc.2016.03.002

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.006

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition and Emotion, 14,* 433–440. http://dx.doi.org/10.1080/026999300402745

Schweppe, J., & Rummer, R. (2014). Attention, working memory, and long-term memory in multimedia learning: An integrated perspective based on process models of working memory. *Educational Psychology Review, 26,* 285–306. http://dx.doi.org/10.1007/s10648-013-9242-2

Shen, L., Wang, M., & Shen, R. (2009). Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment. *Educational Technology & Society, 12,* 176–189.

Shuman, V., & Scherer, K. R. (2014). Concepts and structures of emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 13–35). New York, NY: Routledge.

Sitzmann, T., & Johnson, S. (2014). The paradox of seduction by irrelevant details: How irrelevant information helps and hinders self-regulated learning. *Learning and Individual Differences, 34,* 1–11. http://dx.doi.org/10.1016/j.lindif.2014.05.009

Sukalla, F., Bilandzic, H., Bolls, P. D., & Busselle, R. W. (2015). Embodiment of narrative engagement. *Journal of Media Psychology, 28,* 175–186. http://dx.doi.org/10.1027/1864-1105/a000153

Sung, E., & Mayer, R. E. (2012). When graphics improve liking but not learning from online lessons. *Computers in Human Behavior, 28,* 1618–1625. http://dx.doi.org/10.1016/j.chb.2012.03.026

Sweller, J. (2010). Element interactivity and intrinsic, extraneous and germane cognitive load. *Educational Psychology Review, 22,* 123–138. http://dx.doi.org/10.1007/s10648-010-9128-5

Takahashi, S. (1995). Aesthetic properties of pictorial perception. *Psychological Review, 102,* 671–683. http://dx.doi.org/10.1037/0033-295X.102.4.671

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological Science, 10,* 297–303. http://dx.doi.org/10.1111/1467-9280.00157

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104,* 485–498. http://dx.doi.org/10.1037/a0026609

Van Rompay, T. J., De Vries, P. W., & Van Venrooij, X. G. (2010). More than words: On the importance of picture–text congruence in the online environment. *Journal of Interactive Marketing, 24,* 22–30. http://dx.doi.org/10.1016/j.intmar.2009.10.003

Wang, Y., & Crooks, S. M. (2015). Does the personalization of multimedia instruction influence the effectiveness of decorative graphics during foreign language instruction? *Journal of Technology and Chinese Language Teaching, 6,* 29–38.

Winkielman, P., & Cacioppo, J. T. (2001). Mind at ease puts a smile on the face: Psychophysiological evidence that processing facilitation elicits positive affect. *Journal of Personality and Social Psychology, 81,* 989–1000. http://dx.doi.org/10.1037/0022-3514.81.6.989

Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Mahwah, NJ: Lawrence Erlbaum.

Received September 23, 2016
Revision received April 20, 2017
Accepted April 20, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 250–268

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000221

Animated Pedagogical Agents as Aids in Multimedia Learning: Effects on Eye-Fixations During Learning and Learning Outcomes

Fuxing Wang
Central China Normal University and University of California, Santa Barbara

Wenjing Li
Central China Normal University

Richard E. Mayer
University of California, Santa Barbara

Huashan Liu
Central China Normal University

The goal of the present study is to determine how to incorporate social cues such as gesturing in animated pedagogical agents (PAs) for online multimedia lessons in ways that promote student learning. In 3 experiments, college students learned about synaptic transmission from a multimedia narrated presentation while their eye movements were tracked and subsequently took learning outcome tests. In Experiments 1 and 2, students who had a gesturing PA added to the screen performed significantly better on learning outcome tests of transfer (*ds* = 0.77 and 0.80) and retention (*ds* = 1.16 and 1.00) and spent more time attending to target material based on eye-tracking measures including fixation time (*ds* = 1.53 and 2.27) and number of fixations (*ds* = 1.54 and 1.70). In Experiments 2 and 3, students who learned with a gesturing PA outperformed those who learned with a static PA on transfer (*ds* = 0.72 and 1.02), retention (*ds* = 0.96 and 0.93), fixation time (*ds* = 2.07 and 1.82), and number of fixations (*ds* = 1.64 and 2.99). In Experiment 2, adding a static PA to the screen did not improve performance. In Experiment 3, adding signaling such as color coding did not improve performance for students who received a gesturing PA. Results support the embodiment principle that people learn better from onscreen multimedia lessons when a gesturing PA is added to the screen, and social agency theory, which posits that social cues can prime learners to process the material more actively and develop better learning outcomes.

Educational Impact and Implications Statement

How can we help students learn scientific content that is presented in online multimedia lessons consisting of graphics and narration? Across 3 experiments, students learned better when the screen also included an onscreen character who gestured as she explained the process of neural transmission as compared to identical lessons with no or motionless onscreen characters. Onscreen agents can help create a sense of social partnership that primes learners to try harder to make sense of the material.

Keywords: pedagogical agents, multimedia learning, gesture, signaling, embodiment

Traditional classroom teaching involves learning face-to-face from a teacher who explains important knowledge, often using social cues such as gestures to guide attention and help students understand the content. Recent advances in computer technology, artificial intelligence, and virtual reality technology, allow instructional designers to create vivid onscreen pedagogical agents (PAs) in multimedia learning environments, but research is needed on how to make the PAs as effective as possible in promoting learning. A PA is an image of a character presented on a screen intended to help student learning (Dehn & Van Mulken, 2000; Heidig & Clarebout, 2011; Moreno, 2005; Veletsianos & Russell, 2014). The goal of the present study is to determine how to incorporate

This article was published Online First August 17, 2017.

Fuxing Wang, Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, School of Psychology, Central China Normal University, and Department of Psychological and Brain Sciences, University of California, Santa Barbara; Wenjing Li, Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, School of Psychology, Central China Normal University; Richard E. Mayer, Department of Psychological and Brain Sciences, University of California, Santa Barbara; Huashan Liu, Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, School of Psychology, Central China Normal University.

Wenjing Li is co-first author of this paper.

This research was supported by the National Natural Science Foundation of China Grant (31300864) and Self-Determined Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE (CCNU16A02023) to Fuxing Wang.

Correspondence concerning this article should be addressed to Fuxing Wang or Huashan Liu, School of Psychology, Central China Normal University, 152 Luoyu Street, Wuhan, Hubei, 430079, P. R. China. E-mail: fxwang@mail.ccnu.edu.cn or hsliupsycho@263.net

social cues such as gesturing in animated PAs for online instruction in ways that promote student learning.

Literature Review

During the past 20 years, research on animated PA has progressed through three research questions—can it work, does it work, and how does it work. Early developmental research, in the can-it-work genre, focused on the feasibility of building onscreen agents sometimes with a focus on the user's affective response (Cassell, Sullivan, Prevost, & Churchill, 2000; Johnson, Rickel, & Lester, 2000; Lester et al., 1997; Lester, Towns, & Fitzgerald, 1998). Subsequent experimental research in the does-it-work genre found evidence that students can learn better when computer-based lessons include lifelike onscreen PAs (Dehn & Van Mulken, 2000; Johnson & Lester, 2016; Kim & Baylor, 2016; Moreno, Mayer, Spires, & Lester, 2001; Schroeder & Adesope, 2014; Schroeder, Adesope, & Gilbert, 2013; Veletsianos & Russell, 2014). More focused experimental research in the how-does-it-work genre seeks to pinpoint the conditions under which onscreen agents are most useful such as when they use human-like gestures (Baylor & Kim, 2009; Dunsworth & Atkinson, 2007; Lusk & Atkinson, 2007; Mayer & DaPra, 2012), friendly voice (Atkinson, Mayer, & Merrill, 2005; Mayer & DaPra, 2012), and conversational style based on cognitive theories of learning (Moreno & Mayer, 2000, 2004). The present study fits within the second genre by examining whether adding onscreen agents improves student learning from computer-based lessons (Experiments 1, 2, and 3), and fits within the third genre by examining whether and how adding gesturing to onscreen agents improves student learning from computer-based lessons (Experiments 2 and 3).

First, a primary issue concerns whether adding a PA to an online lesson will improve student learning, which can be called the *pedagogical agent hypothesis*. The literature on the effects of PAs on learning yields mixed results (Heidig & Clarebout, 2011; Schroeder et al., 2013; Wang, Li, Xie, & Liu, 2017), with some studies finding that agents improved learning from multimedia presentations (e.g., Atkinson, 2002; Dunsworth & Atkinson, 2007; Holmes, 2007; Lusk & Atkinson, 2007; Mayer & DaPra, 2012), and other studies finding that agents did not improve learning outcomes (e.g., Bailenson, Swinth, Hoyt, & Persky, 2005; Dirkin, Mishra, & Altermatt, 2005; Frechette & Moreno, 2010; Unal-Colak & Ozan, 2012). A meta-analysis by Schroeder et al. (2013) showed that the presence of PAs produced a small but significant effect on learning ($d = 0.19$). More recently, a meta-analysis by Wang et al. (2017) found that adding a PA to online lessons effectively improved scores on retention tests ($g = 0.19$), transfer tests ($g = 0.39$), and other tests ($g = 0.31$). In Experiments 1 and 2 of the present study, we contribute to the research literature on the PA hypothesis by comparing the effects of learning from an online multimedia lesson that either contains or does not contain an animated PA who gestures while talking.

Second, a more focused issue concerns that features of an onscreen PA promote learning, such as whether having a PA display social cues including gesturing results in better learning than having a PA that does not gesture, which can be called the *social cue hypothesis*. If teachers use social cues (e.g., gesture) in real teaching venues, students perform better in arithmetic lessons (Singer & Goldin-Meadow, 2005) and word learning tasks (McGregor, 2008; McGregor, Rohlfing, Bean, & Marschner, 2009). de Koning and Tabbers (2013) and Fiorella and Mayer (2016) found that even adding a pointing hand in multimedia learning (not including PA's visual image) could improve learning.

There also is preliminary evidence showing that a high embodied PA—including dynamic facial expression, eye gaze, gesture, and body movement—can under certain circumstances promote learning better than a low embodied PA that lacks these feature (Mayer, 2014; Moreno, 2005; Moreno et al., 2001). Studies showed that compared with a static PA, when a PA exhibited social cues such as gesturing and expression, learners performed better (Craig, Twyford, Irigoyen, & Zipp, 2015; Lusk & Atkinson, 2007; Mayer & DaPra, 2012). Based on this, researchers proposed that the positive effects of PA depend on the role of social cues (such as gesturing) rather than the PA's visual image (Clark & Choi, 2005; Mayer, 2014). What's more, they argued that the reason PAs could promote multimedia learning was that the social cue could guide learners' attention (Craig et al., 2015; Johnson, Ozogul, Moreno, & Reisslein, 2013). In Experiments 2 and 3 of the present study we seek to test the social cue hypothesis by focusing on whether students learn better from a multimedia lesson that contains a PA who gestures versus a PA that does not, including using eye tracking to record cognitive processing of participants during learning.

Third, a related issue examined in Experiment 2 concerns whether adding the PA's visual image (without social cues such as gesturing) can improve learning, which can be called the *image hypothesis*. The PA's visual image mainly refers to presenting a static character on the computer screen. In previous studies, a variety of images of agents have been used, for example, humanoid animals (Lusk & Atkinson, 2007; Yılmaz & Kılıç-Çakmak, 2012), cartoon characters (Johnson, Ozogul, & Reisslein, 2015; Yung & Paas, 2015), and human or human-like images (Moreno & Flowerday, 2006; Unal-Colak & Ozan, 2012). Some studies found that static PA images improved learning outcomes (Frechette & Moreno, 2010; Jin, 2010; Yılmaz & Kılıç-Çakmak, 2012), but a recent meta-analysis involving 14 experimental comparisons found a negligible median effect size of $d = 0.20$ (Mayer, 2014).

Fourth, it is possible any positive effects caused by adding a gesturing PA can be attributed to the fact that the PA uses pointing gestures that signal where to look in the slide. If the PA's gestures serve mainly as signaling cues, then we should be able to replace an onscreen agent with signals such as arrows pointing where to look in sync with the narration. We refer to this idea as the *signaling hypothesis*. Many studies have found that signaling or cuing (e.g., with arrows, color coding, highlighting, or spotlights) can effectively guide learner's attention (Boucheix & Lowe, 2010; de Koning, Tabbers, Rikers, & Paas, 2007; Ozcelik, Karakus, Kursun, & Cagiltay, 2009; van Gog, 2014; Wang, Duan, & Zhou, 2013) and improve performance in multimedia learning (Boucheix & Lowe, 2010; Mautone & Mayer, 2007; Ozcelik et al., 2009). For example, de Koning, Tabbers, Rikers, and Paas (2010a) used eye tracking in their study and found that participants had a higher fixation count and longer fixation time on the cued area. de Koning, Tabbers, Rikers, and Paas (2010b) found that when learning about the human cardiovascular system, learners in the cued animation had higher retention and transfer scores than those in the uncued animation. Therefore, some researchers proposed that we can replace complicated social cues with simple physical cues (Choi & Clark, 2006; Clark & Choi, 2005). In Experiment 3 of the present study, we test the signaling hypothesis by comparing the test performance of students who learn from a slideshow lesson that contains a PA who exhibits social cues (i.e., gesturing) versus learning

from a slideshow lesson that contains signaling in the form of color coding in which elements in the illustration are highlighted in red when discussed by the PA.

Theory and Predictions

How does the PA affect learning processes and outcomes? There are competing theories to explain the learning mechanism underlying any effects of PAs, with social agency theory and theories based on embodiment making the case for using highly embodied PAs and cognitive load theory and theories based on seductive details making the case against them.

One theoretical approach is based on the idea that social cues in the learning environment can affect the learner's motivation to engage in appropriate cognitive processing of the material. Specifically, according to social agency theory, the presentation of social cues (such as the PA's conversational style, voice, image, and gestures) helps build a feeling of social partnership in the learner, causing the learner to exert more effort to engage in deep cognitive processing during learning, which is more likely to lead to meaningful leaning outcomes (Mayer, 2014; Mayer & DaPra, 2012). The first step in the process occurs when social cues in the instructional message create a sense of social presence in the learner. Moreno and Mayer (2004) have shown that learning is improved by social cues that create a sense of social presence—a feeling of interacting with another human being—but learning is not necessary improved by physical cues that create a sense of physical presence—a feeling of being in a realistic environment.

The second step in the process occurs when learners exert more effort to make sense of the material because they want to understand what a social partner is saying. Based on Grice's (1975) classic *cooperation principle* of human-to-human communication, there is an implicit contract between speaker and listener in a conversation such that the listener will try to understand what the speaker is saying and the speaker will try to be clear for the listener. In short, when people feel a sense of social partnership with the on-screen agent, they will try harder to understand what a speaker is saying, consistent with this implicit contract. Consistent with social agency theory, Reeves and Nass (1996) and Nass and Brave (2005) have shown how people easily can accept a computer as a social partner as long as appropriate social cues are present such as human-like voice and onscreen image, which they refer to as *the media equation*.

In the present study we measured cognitive engagement during learning through five eye-tracking measures: fixation time (total time spent with eyes fixated on relevant elements), fixation count (number of fixations on relevant elements), average fixation (mean time per fixation on relevant elements), first fixation duration (time spent on first fixation on each relevant element), and glances count (number of saccades to relevant elements). We interpret higher mean scores on these variables as indicating deeper cognitive processing during learning aimed at attending to relevant material (e.g., fixation count, fixation count) and organizing and integrating it (e.g., average fixation, first fixation duration, and glances count). The use of eye-tracking measures represents an attempt to move beyond self-reports of cognitive effort typically used in instructional studies, allowing for a direct behavioral measure of cognitive activity during learning.

The third step occurs when learners who exerted more effort to process the incoming material deeply during learning also show better performance on subsequent tests of what they have learned. Much of the foundational research on social agency theory showing learning improvements involves scientific, technical, engineering, or mathematical (STEM) material such as explanations of how solar cells work (Mayer & DaPra, 2012), how lightning storms develop (Mayer, Sobko, & Mautone, 2003; Moreno & Mayer, 2000), how plants grow (Moreno & Mayer, 2000, 2004), how the human respiratory system works (Mayer, Fennell, Farmer, & Campbell, 2004), how the human cardiovascular system works (Dunsworth & Atkinson, 2007), astronomy (Frechette & Moreno, 2010), chemistry (McLaren, DeLeeuw, & Mayer, 2011a, 2011b), electrical circuits (Moreno, Reislein, & Ozogul, 2010), mathematics word problems (Atkinson et al., 2005; Lusk & Atkinson, 2007), and word processing software (Baylor & Kim, 2009). Perhaps, social cues are useful with STEM material that might not otherwise have a strong social connection to learners.

In the present study, we measured meaningful learning outcomes through three learning outcome posttests: retention, transfer, and matching. The retention test asked learners to write down the key steps in the process described in the lesson and, thus, tapped the degree to which learners attended to relevant material in the lesson. The transfer asked learners to answer five open-ended questions that required using the material in a new way and, thus, tapped the degree to which learners understood the material through having organized and integrated it. The matching test involved fill in the name of elements in an illustration and, thus, tapped the degree to which students integrated words and graphics, which is a key step in understanding the material.

Based on social agency theory we can offer two major hypotheses concerning the effects of adding a gesturing agent to an online multimedia lesson. The first hypothesis focuses on effects on the three learning outcome measures obtained through administering posttest after instruction and the second one focuses on effects on the five learning process measures obtained through eye-tracking during learning.

> *Hypothesis 1:* Students who learn with a gesturing PA will outperform students who learn without a gesturing PA on retention test, transfer test, and matching tests, which are intended to measure meaningful learning outcomes.

> *Hypothesis 2:* Students who learn with a gesturing PA will outperform students who learn without a gesturing PA on eye-tracking measures including fixation time, fixation count, average fixation, first fixation duration, and glances count, which are intended to measure meaningful learning processes.

In short, based on social agency theory, we can predict that adding a gesturing PA to the screen should result in deeper (or more meaningful) learning processes and outcomes than not having a gesturing PA, thereby confirming the PA hypothesis. This prediction is tested in two experiments (Experiments 1 and 2). In addition, based on social agency theory, we offer two complementary hypotheses concerning the role of embodiment and the role of the signaling.

> *Hypothesis 3:* Students who learn with high embodied PAs (i.e., PAs who gesture) should display better learning outcomes (as measured by scores on three posttests) and better learning processes (as measured by the five eye-tracking measures) than low embodied PAs, consistent with the social cues hypothesis and inconsistent with the image hypothesis. In

short, this prediction is based on the idea that a high level of embodiment is the active ingredient that causes deeper learning with PAs. This prediction is addressed in Experiment 2.

Hypothesis 4: Students who learn with gesturing PAs will display improvements in learning outcomes and processes that go beyond the effects of simply adding physical cueing, in contrast to the signaling hypothesis. Further, adding physical cueing should improve learning processes and outcomes for students who learn without gesturing PAs, but should not help for students who learning with gesturing PAs. In short, this prediction is based on the idea the benefits of gesturing PAs are not attributable solely to the fact that the agent's pointing gestures are a form of signaling. This prediction is addressed in Experiment 3.

On the other hand, according to cognitive load theory (Paas, Renkl, & Sweller, 2003; Sweller, Ayres, & Kalyuga, 2011) and theories of seductive details (Harp & Mayer, 1998), a PA can be a seductive detail—an interesting but irrelevant part of a lesson—that increases extraneous cognitive load—cognitive processing that does not support the instructional objective—thereby leaving less remaining cognitive capacity for learning the essential material. During learning, the PA could attract the learner's attention thereby leaving less cognitive capacity for attending to and mentally representing the important content of learning materials (Mayer, 2009; Mayer & Fiorella, 2014). In the organization and integration phases of information processing, learners need to simultaneously process the PA and learning materials, which may cause high cognitive load and thereby impede the learning outcomes (Baylor & Ryu, 2003; Moreno et al., 2001). If PAs serve as seductive details, in contrast to our four hypotheses, we would predict no support for the PA hypothesis or social cue hypothesis and positive support for a version of the signaling hypothesis in which learners benefit more from physical signals such as highlighting with color than from gesturing PAs who use pointing.

Consistent with the seductive details theory, previous research has shown that adding interesting but irrelevant photos or video to science lessons tends to diminish learning (Harp & Mayer, 1998; Mayer, Heiser, & Lonn, 2001). Also consistent with the seductive details theory, a recent review (Mayer, 2014) found that adding a motionless onscreen character to an online science lesson generally does not improve learning and in some cases diminishes learning (Mayer & DaPra, 2012). A motionless PA can be distracting because it does not display the kind of humanlike motion that learners expect. However, we attempted to eliminate this potential in the present study by using embodied PAs, that is, PAs that used human-like gesture and motion.

A major advance in addressing these questions in the present study involves the use of multileveled learning tests to measure learning outcomes (e.g., retention, transfer, and matching tests) coupled with eye-tracking measures to measure learning processes mainly tapping the degree to which learners attend to relevant portions of the illustration (e.g., number of fixations and fixation time) and organize and integrate the material (e.g., average duration, first fixation duration, and glance count). We operationalize our measure of learning outcome by examining scores on learning tests given after the lesson with higher scores indicating deeper learning. We include multileveled tests that tap both amount remembered and ability to use the material in new situations. We operationalize our measure of learning process by examining scores on eye-tracking measures tapping attention to the area of interest (AOI) being described by the PA with higher scores indicating more attention to relevant portions of the graphic. We include eye-tracking techniques that record the individual's attention distribution in real time (Rayner, 1998, 2009) and thereby investigate the process of knowledge extraction and attention distribution in multimedia learning (Hyönä, 2010; Mayer, 2010).

Experiment 1

Researchers have found mixed results of the effects of PAs on learning, in which PAs help learning in some situations but not in others (Heidig & Clarebout, 2011; Schroeder & Adesope, 2014; Schroeder et al., 2013; Veletsianos & Russell, 2014); therefore, we wanted to pinpoint gesturing as a feature of PAs that we hypothesize makes them effective. As a first step, in Experiment 1, we explore whether adding an appealing PA who points to relevant portions of an illustration could produce superior learning outcomes (as measured by retention and transfer tests) and prime efficient cognitive processing during learning (as measured by eye-tracking measures).

Method

Participants and design. The participants were 51 undergraduates recruited from a university in central China. Their mean age was 20.4 years ($SD = 2.4$) and 39 of them were women. The experiment used a between-subjects design with 25 participants in the PA group, and 26 in the no PA group (No PA). All participants had normal or corrected-to-normal vision, and Chinese was their native language. They were majoring in Psychology (14), English (12), Education (9), Politics (7), Government Management (2), Geography (1), Biology (1), Mathematics (1), Tourism Management (1), Journalism (1), Chinese (1), and Music (1). There was no significant difference between the PA group and the No PA group on prior knowledge based on a pretest, $t(49) = 0.35, p > .05$, mean age, $t(49) = -1.87, p > .05$, and proportion of men and women, $\chi^2(1) = 0.34, p > .05$. The study was approved by the ethics committee of the university where the study was conducted, and the study followed standards for ethical treatment of human subjects.

Materials and apparatus. The materials consisted of two versions of a computer-based multimedia lesson on synaptic transmission, a pretest, a retention test, a transfer test, and a matching test.[1] All materials were in Chinese.

Multimedia lessons. As exemplified in Figure 1, the two narrated multimedia lessons described the process of chemical

[1] In each of the three experiments we administered a postquestionnaire after the lesson. The postquestionnaire asked students to rate four subjective questions intended to measure learning perceptions concerning mental effort ("How much effort did you put in learning process?"), motivation ("How much would you like to learn other learning materials in this way?"), interest ("How interesting was the learning material?"), and social partnership ("How much did you feel that there was a real person talking to you?"). The first three items were on a 9-point scale ranging from 1 (*very little*) to 9 (*very much*); the fourth was on a 5-point scale ranging from 1 (*very little*) to 5 (*very much*). However, the postquestionnaire generally proved to be insensitive to differences among the groups, so we do not report the results here.

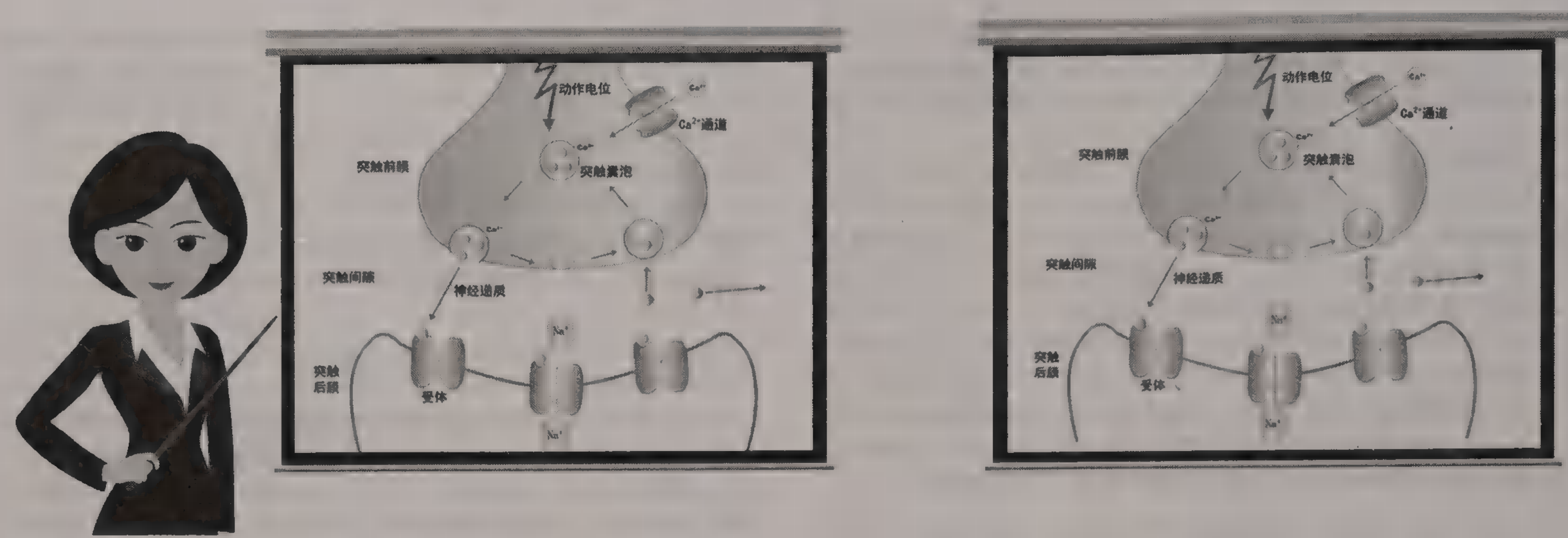

Figure 1. Example frames from animations in Experiment 1. Left is pedagogical agent (PA) version, and right is no PA version. See the online article for the color version of this figure.

synaptic transmission, either with or without an on-screen agent standing to the left of the on-screen illustration. The lessons focused on how signals are transmitted across neurons in the nervous system and the role of action potentials, calcium ions, sodium ions, and neurotransmitters. Both versions included narration in a young female voice and an illustration that showed the parts of neurons involved in synaptic transmission. In the PA version there was an animated female agent standing to the left of the illustration who used posture, eye gaze, and pointing gesture (with a handheld pointer) to direct attention to the relevant parts of the illustration as they were discussed in the narration. The No PA version was same as the PA version except there was not an onscreen agent. Both multimedia lessons were created using Flash CS6 software with the screen size of 1,680 × 1,050 pixels and lasted 128 s.

Pretest. The pretest solicited demographic information (e.g., gender, age, major, and educational level) and included 10 multiple-choice questions about chemical synaptic transmission and four subjective rating statements. An example question was, "Where is the neurotransmitter stored before it was released?" Each question had four options, and only one was the correct answer. Two points were awarded for each correct answer. An example rating statement is, "How much do you know about chemical synapses?" The participants needed to mark a 5-point scale ranging from 0 (*very little*) to 4 (*very much*). A prior knowledge score was computed by summing the number of points on the multiple choice items with the number of points on the rating items, yielding a score that could range from 0 to 31.

Learning outcome tests. The learning outcome tests consisted of a retention test, transfer test, and matching test. The retention test required participants to write down the process of chemical synaptic transmission in detail according to what they had learned. One point was awarded for each idea unit representing a key point, regardless of specific wording. A maximum of 22 points could be achieved (see Appendix for a list of the 22 idea units). The transfer test consisted of five open-ended questions examining to what extent learners applied the learning knowledge to novel problems (e.g., "What factors can affect the process of chemical synaptic transmission?"). One point was awarded for each acceptable answer, regardless of specific wording. The total possible score was 17 points. Two raters scored the retention and transfer tests independently and the average of them was used as the learner's final score. Interrater reliability was $r = .98$ ($p < .001$) for the retention test and $r = .98$ ($p < .001$) for the transfer test. The matching test included the same illustration as in the study phase with pointers to each of seven elements but without the labels of the elements. Participants were asked to fill in the names of all elements. One point was awarded for each correct answer. The total possible score was seven points. The Cronbach's α for the three tests was 0.79, which is an acceptable level.

Apparatus. A SMI RED 250 Desktop eye-tracker (SensoMotoric Instruments, Germany) was used to record the eye movement data. The eye-tracker operates at a sampling rate of 250 Hz and has a spatial resolution of less than 0.1°. The computer screen for displaying the animation was positioned 70 cm from the participant with of 1,680 × 1,050 pixels resolution. The fixation filtering threshold was set at 100 ms.

Procedure. Participants were randomly assigned to the PA or No PA group and tested individually. First, the participant completed the pretest test at his or her own pace. Second, the participant was seated in front of the eye tracking monitor, and the experimenter started to calibrate. After that, the participant read instructions for the experiment. Once the participant understood the instructions, the multimedia lesson was presented, which lasted 128 s. After viewing the lesson, the participant completed the matching test, retention test, and transfer test in that order at his or her own pace. The entire experiment took about 25 min. We adhered to guidelines for ethical treatment of human subjects.

Results and Discussion

The results included learning posttest scores and eye-tracking data.

Posttest scores: Does adding a PA to an online multimedia lesson improve student learning? According to the PA agent hypothesis as stated in Hypothesis 1, adding an appealing and gesturing onscreen agent to a multimedia lesson should result in better learning as reflected in higher scores on tests of learning

outcome. Table 1 shows the mean scores (and *SDs*) for the PA and No PA groups on the retention test, transfer test, and matching test. To investigate the PA's effect on multimedia learning outcomes, we conducted analysis of covariances (ANCOVAs) with pretest score as a covariate. Participants in the PA group had higher scores than those in the No PA group on the retention test, $F(1, 48) = 18.00$, $p < .001$, $\eta_p^2 = .27$, and the transfer test, $F(1, 48) = 9.91$, $p < .01$, $\eta_p^2 = .17$. The difference between the two groups on the matching test did not reach significance, $F(1, 48) = 3.99$, $p = .052$, $\eta_p^2 = .08$. The same pattern of results was found when we conducted *t* tests except the difference between the groups on the matching test was statistically significant. As predicted by social agency theory, adding a human-like agent to online lessons improved the memorization and understanding of the presented materials, which also is consistent with previous studies (Frechette & Moreno, 2010; Mayer & DaPra, 2012; Moreno et al., 2010) in which participants performed better in the PA group than the No PA group. Overall, these results provide support for the predictions of the social agency theory. In contrast, these results do not support the idea that adding a gesturing onscreen agent serves as a seductive detail that distracts the learner.

Eye-tracking measures: Does adding a PA to an online multimedia lesson affect eye fixations during learning? According to the PA hypothesis, as stated in Hypothesis 2, adding an appealing onscreen agent who points to relevant material on the accompanying slides should direct the learner's attention to the relevant material and encourage organizing and integrating of the material. To analyze the eye-tracking data we defined 12 AOIs corresponding to each of the main portions of the illustration, as shown in Figure 2. According to social agency hypothesis, adding an appealing onscreen agent who points to relevant elements in the diagram should result in better performance on eye-tracking measures intended to measure efficient cognitive processing. Table 2 shows the mean (and *SD*) on each of six eye-tracking measures intended to indicate efficient cognitive processing:

Fixation time—the number of seconds the student fixates on relevant elements (e.g., the elements being described in the narrative).

Fixation count—the number of times the student fixates on the relevant elements.

Average fixation—sum of fixation time on relevant elements divided by number of fixations on relevant elements.

First fixation duration—the number of seconds the student first fixates on a relevant element.

Table 1
Mean Scores on Learning Posttests and SDs for Two Groups in Experiment 1

Dependent variables	PA *M*	PA *SD*	No PA *M*	No PA *SD*	*df*	*F*	*p*	η_p^2
Retention test	7.57	2.78	4.50	2.60	48	18.00	<.001	.27
Transfer test	5.10	2.24	3.41	2.20	48	9.91	.003	.17
Matching test	5.42	1.62	4.48	2.25	48	3.99	.052	.08

Note. PA = pedagogical agent.

Glances count—the number of times the student saccades to the relevant elements from outside.

To investigate the PA's effect on cognitive processing during learning, we conducted ANCOVAs comparing the two groups on each eye-tracking measure with pretest score as a covariate. As predicted by social agency theory, Table 2 shows that the PA group differed significantly from the No PA group on each of the five eye-tracking measures—indicating that PA group had more fixation time on the target material, more fixations on the target material, longer fixation duration, longer first fixation, and more glances. These results show that adding a human-like agent in multimedia learning did not reduce learner's attention to the learning content. On the contrary, it increased the participants' fixation time and number of fixations on relevant portions of the learning materials. Li et al. (2016) also found that adding an agent did not reduce learners' attention to learning content. These results together provided support for social agency theory and do not support seductive details theory.

Previous studies found that when the agent was accompanied by social cues, learners had better performance (de Koning & Tabbers, 2013). However, Louwerse, Graesser, Mcnamara, and Lu (2009) found that when an onscreen agent was added to a computer-based lesson, learners paid more attention to the agent, rather than the learning content. A possible explanation for this inconsistency was a difference in instructions and context. In Louwerse et al. (2009), learners looked at the screen freely, which caused learners to lack clear learning objectives in the learning process and learners' attention could easy to be attracted by the agent. In contrast, in the present experiment and Li et al. (2016), participants were told that there would be a test after learning, which caused learners to put more attention on the learning materials. Second, in this experiment and Li et al. (2016), the agent was accompanied by gesture, gaze and other social cues, but in Louwerse et al. (2009), the agent was static, not including any social cues.

Eye-tracking measures: Do students look at the PA less over the course of an online multimedia lesson? Although looking at the onscreen agent at the start of the lesson may help build a social partnership, continuing to look at the onscreen agent throughout the lesson is an inefficient cognitive strategy that could lead to less processing of the relevant material in the illustration. To investigate changes in the learners' pattern of attention allocation over the time course of the lesson, we created two AOIs as shown in Figure 3, the agent and the relevant portion of the illustration (i.e., the elements mentioned in the narration). To explore fixation patterns as a function of time, we divided the multimedia lesson into three time periods: introduction to the topic (0-8 s), the first half of the content (9-68 s), and the second half of the content (69-128 s), and then we performed two-way repeated ANOVA with AOI and time period as factors.

Figure 4 shows the relative fixation time (as a percentage) for each AOI in each time period. As can be seen in Table 3, the main effect of AOI was significant, $F(1, 24) = 352.96$, $p < .001$, $\eta_p^2 = .94$, with students spending a significantly greater proportion of time looking at the illustration ($M = 64.27$, $SD = 2.30$) than the on-screen agent ($M = 9.41$, $SD = 1.18$), and the main effect of time period was also significant, $F(2, 48) = 11.57$, $p < .001$, $\eta_p^2 = .33$. Post hoc least significant difference (LSD) tests (with $p < .05$)

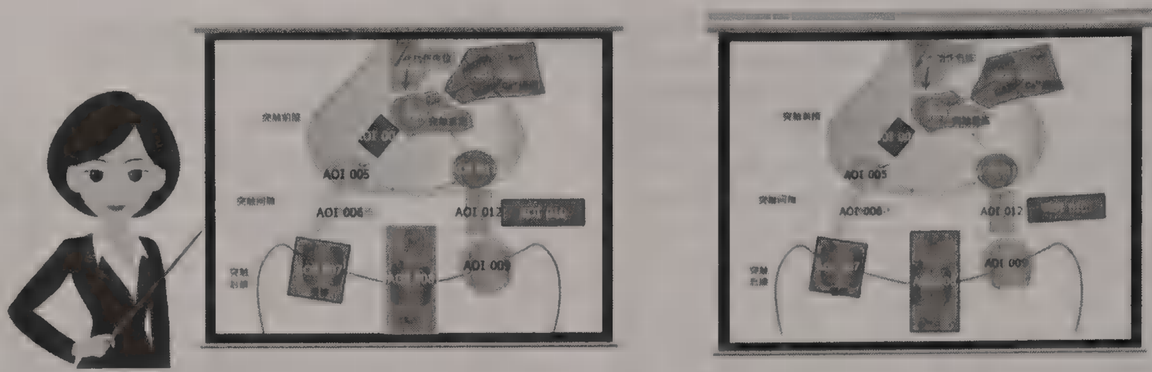

Figure 2. Twelve areas of interest (AOIs) for pedagogical agent (PA) group (left) and no PA group (right) overlapped on one picture for presentation. See the online article for the color version of this figure.

showed that students spent a greater proportion of time at Stage 2 (M = 38.58, SD = 1.07) than Stage 3 (M = 37.26, SD = 1.11) and Stage 1 (M = 34.67, SD = 1.39). There was a significant interaction between AOI and time period, $F(2, 48) = 73.10$, $p < .001$, $\eta_p^2 = .75$,in which the relative fixation time on the agent in Stage 1 was greater than in later stages, $F(2, 48) = 42.61$, $p < .001$; but the relative fixation time on the illustration in later stages was greater than in Stage 1, $F(2, 48) = 72.21$, $p < .001$.

These results show that learners in the PA group do not continuously look at the onscreen agent during the entire lesson, but spend very little time looking at the on-screen agent, and mainly in the first 8 s of the lesson. Clearly, the PA is not serving as a seductive detail that dominates the learner's visual attention throughout the lesson. A possible explanation is that at the start of the lesson, the PA is a new stimulus that attracts some small portion of the learners' attention, but as the lesson progresses, learners become familiar with the agent, leading to reducing even that small amount of attention given to agent and increasing attention to learning content in the illustration.

How do learning outcome measures correlate with eye-tracking measures? To further explore the relationship between learning processing and learning outcomes, a correlation analysis was conducted among the three measures of learning outcome (i.e., retention test score, transfer test score, and matching test score) and the five eye-tracking measures. If efficiency of cognitive processing (as indicated by eye-tracking measures) is related to learning outcomes (as measured by retention, transfer, and matching tests) as predicted by agency theory, we expect significant correlations between those measures. Table 4 shows that all 15 of these key correlations between a learning outcome measure and an eye-tracking measure are in the predicted direction, and 8 of the 15 correlations are statistically significant at the $p < .05$ level. This pattern of results is consistent with the idea that efficient visual processing during learning was related to superior scores on tests of learning outcome. The eye-tracking results add a new line of support for social agency theory, which posits that adding a gesturing PA will cause deeper cognitive processing during learning (as reflected in the eye-tracking measures, such as greater fixation time on the target material) leading to better learning outcomes (as reflected in better posttest scores).

Experiment 2

Experiment 1 revealed positive effects on learning processes and learning outcomes of adding an onscreen PA who gestured during an online slideshow lesson as compared with presenting an onscreen multimedia lesson without a PA. However, it is not possible to tell which aspects of the PA were the active ingredients in improving learning—the presence of the agent's image on the screen (i.e., image effect), the presence of an agent who engaged in human-like gesturing (i.e., gesture effect), using a pointer to identify specific spots on the illustration (i.e., signaling effect), or a combination of all three elements. In the present study, we sought to disentangle the effects of these three kinds of cues in the PA by comparing groups that received a gesturing agent as in Experiment 1 (with all 3 elements), no PA as in Experiment 1 (with none of the elements), a group that had a nongesturing agent (with only the image element), and a group that had a pointer but no agent (with only the signaling element).

Table 2
Mean Scores and SDs on Eye-Tracking Measures for Two Groups in Experiment 1

	PA		No PA					
Dependent variables	M	SD	M	SD	df	F	p	η_p^2
Fixation time (ms)	49,125	12,486	33,323	8,216	48	28.84	<.001	.38
Fixation count	138.60	31.75	98.54	20.46	48	28.66	<.001	.37
Average fixation (ms)	365	118	279	65	48	10.54	.002	.18
First fixation duration (ms)	308	85	258	67	48	5.28	.026	.10
Glances count	100.96	23.03	75.81	16.73	48	20.01	<.001	.29

Note. PA = pedagogical agent.

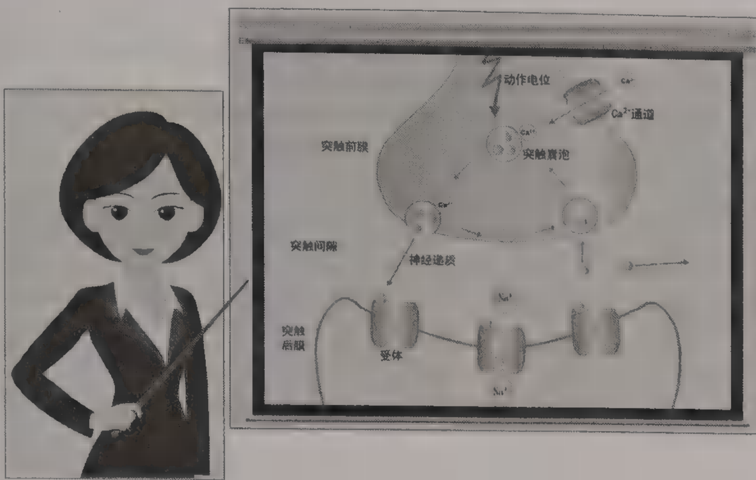

Figure 3. Two areas of interest (AOIs) for agent and illustration. See the online article for the color version of this figure.

Method

Participants and design. There were 109 undergraduates who were recruited from a university in central China. The mean age was 19.9 years (SD = 1.9) and 96 of the participants were women. This experiment used a 2 × 2 between-subjects design with PA (PA vs. No PA) and gesture (gesture vs. no gesture) as factors. The participants were randomly assigned to four groups: 27 in the PA with gesture (PA-gesture) group, 27 in the PA without gesture (PA-no gesture) group, 26 in the no PA with gesture (No PA-gesture) group, and 29 in the no PA without gesture (No PA-no-gesture) group. All participants had normal or corrected-to-normal vision and Chinese was their native language. There was no significant difference among the groups on mean age, $F(3, 105) = 1.17$, $p > .05$, pretest score, $F(3, 105) = 0.42$, $p > .05$, and proportion of men and women, $\chi^2(3) = 0.84$, $p > .05$. They were majoring in Psychology (25), Chemistry (19), Education (17), Mathematics (15), Computer (7), Physics (6), English (4), Chinese (4), Politics (3), History (2), Law (2), Geography (2), Management (1), Japanese (1), and Biology (1).

Materials and apparatus. The apparatus was the same as in Experiment 1. The instructional materials were the same as in Experiment 1 except there were four multimedia lessons that described the process of chemical synaptic transmission, either with or without an on-screen agent standing at the left of the on-screen illustrations and either with or without gesturing that included hand pointing with a pointer to relevant elements in the illustration. As shown in Figure 5, the PA-gesture version was the same as the PA version in Experiment 1 and the No PA-no-gesture version was the same as the No PA version in Experiment 1. The PA-no-gesture version was the same as the PA with gesture version except that the agent was static without any gesturing (including no pointing to the illustration). The No PA-gesture version showed only a handheld pointer that identified relevant parts within the illustration (with no image of the PA's body on the screen).

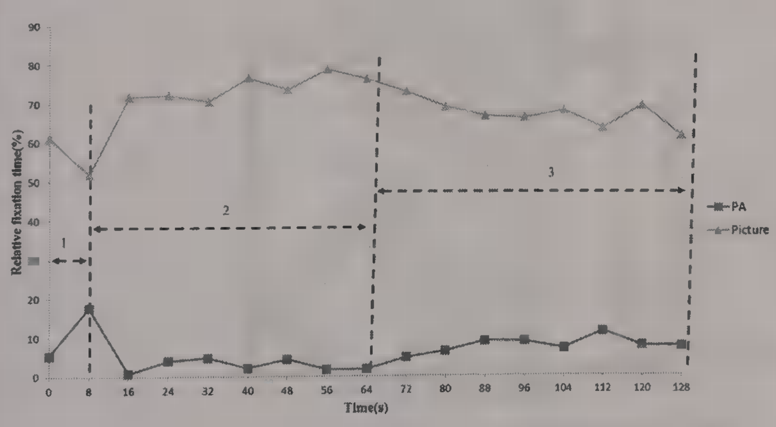

Figure 4. The time process graph of the agent and picture. See the online article for the color version of this figure.

Table 3
The Relative Fixation Time (%) on the Picture and Agent in Experiment 1

Stage	Time	Picture M	Picture SD	Agent M	Agent SD
1	0–8 s	51.82	14.08	17.52	10.08
2	9–68 s	74.47	11.07	2.69	3.19
3	69–128 s	66.51	12.90	8.01	7.72

The pretest questionnaire, matching test, retention test, and transfer test were the same as in Experiment 1. Interrater reliability was $r = .96$ ($p < .001$) for the retention test and $r = .98$ ($p < .001$) for the transfer test. The Cronbach's α for these three tests was 0.74.

Procedure. The procedure was the same as Experiment 1, except that participants were randomly assigned to four groups.

Results and Discussion

Posttest scores: How does the PA's level of embodiment affect learning performance? If the PA effect is caused by the combination of image, gesturing, and signaling (i.e., social cue hypothesis as reflected in Hypothesis 3), then the PA-gesture group should outperform the control group as in Experiment 1, and the other two groups should not. If the PA effect is caused mainly by having the agent's image on the screen, then the PA-no-gesture group should outperform the control group (No PA-no-gesture) and be equivalent to the PA-gesture group (i.e., image hypothesis). If the PA effect is caused mainly by the signaling cue of the pointer identifying where to look on the illustration (i.e., signaling hypothesis), then the No PA-gesture group should outperform the control group (No PA-no-gesture) and be equivalent to the PA-gesture group. Table 5 shows the means (and SDs) of the four groups on the retention test, transfer test, and matching test. To disentangle the effects of the features of the PA group, we performed two-way ANCOVAs with image (PA vs. No PA) and gesture (gesture vs. no gesture) as factors, and pretest scores as a covariate. The same pattern of significant differences was obtained when we conducted an ANOVA.

For the retention test, results showed that the main effect of image was significant, $F(1, 104) = 4.45$, $p < .05$, $\eta_p^2 = .04$, with participants in PA groups (M = 7.46, SD = 3.06) outperforming those in No PA groups (M = 6.22, SD = 2.57), and the main effect of gesture was also significant, $F(1, 104) = 16.31$, $p < .001$, $\eta_p^2 = .14$, with participants in gesture groups (M = 7.75, SD = 3.04) outperforming those in the no gesture groups (M = 5.97, SD = 2.44). These main effects were qualified by a significant interac-

Table 4
The Correlation Between Learning Outcome Scores and Eye-Tracking Measures in Experiment 1

Dependent variables	Retention test	Transfer test	Matching test	Fixation time	Fixation count	Average fixation	First fixation duration
Retention test							
Transfer test	.53**						
Matching test	.57**	.67**					
Fixation time	.41**	.31*	.49**				
Fixation count	.18	.35*	.16	.50**			
Average fixation	.29*	.14	.37**	.63**	.08		
First fixation duration	.16	.07	.36*	.62**	.06	.66**	
Glances count	.14	.32*	.15	.37**	.95**	−.12	.05

* $p < .05$. ** $p < .01$.

tion between image and gesture, $F(1, 104) = 4.65$, $p < .05$, $\eta_p^2 = .04$, in which the PA group significantly outscored the no PA group when there was gesture, $F(1, 105) = 8.43$, $p < .01$), but not when there was no gesture, $F < 1$; and the gesture group significantly outperformed the no gesture group when there was an agent on the screen, $F(1, 105) = 13.11$, $p < .001$; but not when there was no agent, $F < 1$.

For the transfer test, two-way ANCOVA identified a significant main effect of image, $F(1, 104) = 5.26$, $p < .05$, $\eta_p^2 = .05$, with participants in the PA groups ($M = 5.66$, $SD = 1.91$) outperforming those in the No PA groups ($M = 4.70$, $SD = 2.06$), and a significant main effect of gesture, $F(1, 104) = 5.35$, $p < .001$, $\eta_p^2 = .05$, with participants in the gesture groups ($M = 5.56$, $SD = 2.10$) outperforming those in no gesture groups ($M = 4.81$, $SD = 1.93$). The interaction between image and gesture did not reach significance, $F(1, 104) = 3.71$, $p = .057$, $\eta_p^2 = .03$; however, the PA group significantly outperformed the No PA group when there was gesture, $F(1, 105) = 8.13$, $p < .01$, but not when there was not gesture ($F < 1$); and the gesture group outperformed no gesture group when there was agent on the screen, $F(1, 105) = 5.94$, $p < .05$, but not when there was no agent ($F < 1$).

For the matching test, the main effect of image was not significant, $F(1, 104) = 1.42$, $p > .05$, the main effect of gesture was also not significant, $F(1, 104) = 1.96$, $p > .05$, and the interaction between image and gesture was significant, $F(1, 104) = 4.67$, $p < .05$, $\eta_p^2 = .04$. The analysis of simple main effects indicated that the PA group significantly outscored the no PA group when there was gesture, $F(1, 105) = 5.25$, $p < .05$, but not when there was no gesture, $F < 1$; and the gesture group outperformed no gesture group when there was an agent in the screen, $F(1, 105) = 3.88$, $p = .052$, but not when there was no agent, $F < 1$.

Overall, across all three measures of learning outcome, there is a consistent pattern in which the PA-gesture group strongly outperformed the control group (as in Experiment 1) and the other

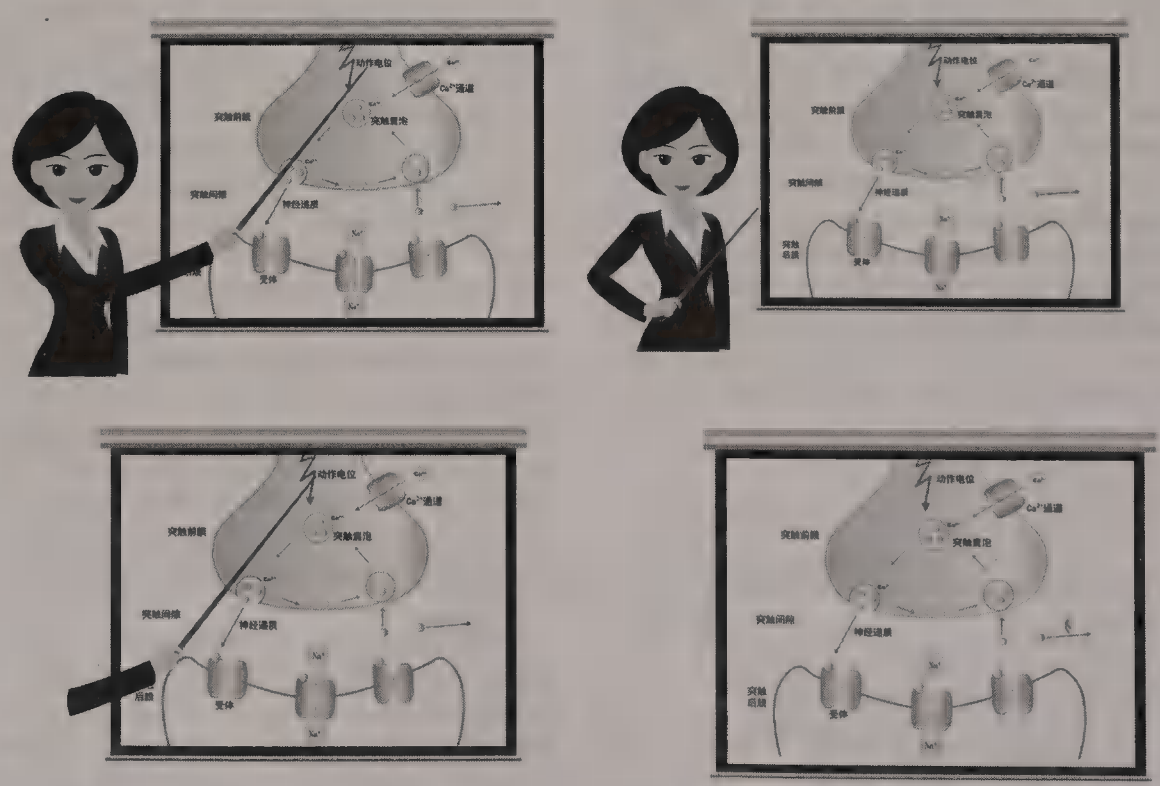

Figure 5. Example frame from animations in Experiment 2. Upper left is pedagogical agent (PA) with gesture version, upper right is PA with no gesture version, lower left is no-PA with gesture, and lower right is no-PA with no-gesture. See the online article for the color version of this figure.

Table 5
Mean Scores on Learning Posttests and SDs for Two Groups in Experiment 2

Dependent variables	PA				No PA			
	Gesture		No gesture		Gesture		No gesture	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Retention test	8.78	3.27	6.15	2.19	6.68	2.40	5.80	2.68
Transfer test	6.31	2.01	5.01	1.60	4.79	1.92	4.62	2.21
Matching test	5.98	1.41	5.07	1.74	4.92	1.75	5.09	1.83

Note. PA = pedagogical agent.

groups do not. This pattern of results supports the social cue hypothesis that PAs are effective when they are highly embodied—that is when there is an image of the PA's body on the screen that gestures and points. There is not convincing evidence for the image hypothesis, because having the PA's image on the screen is not helpful when the PA does not gesture and point. Similarly, there is not convincing evidence for the signaling hypothesis because pointing to elements on the screen was not helpful when the PA's image was not on the screen.

These results are consistent with earlier findings by Mayer and DaPra (2012) in which learners perform better learning from a lesson with a high embodied agent who displays gestures and facial expression than a low embodied agent who stands still. Similarly, some researchers suggest that PAs promote learning because of gesture's role, not the role of agent's image on screen (Choi & Clark, 2006). The main contribution of Experiment 2 is that PAs are most effective when they involve all three features-image, gesturing, and signaling-rather than just having the PA's image or just providing signaling.

Eye-tracking measures: How does the PA's level of embodiment affect students' eye fixations during learning? If visual signaling (i.e., by a PA who moves a pointer or simply by showing a pointer without a PA) guides the learner's attention to the relevant parts of the illustration as they are mentioned, then learners should score higher on eye-tracking measures when there is a pointer that points to the relevant portions of the illustration, as in the PA-gesture group or the No PA-gesture group (de Koning et al., 2010a; Hyönä, 2010). As in Experiment 1, we first defined 13 components as AOIs, as shown in Figure 6. To assess the effectiveness of the PA's cues, we, therefore, created temporary AOIs corresponding to each cued component to find out whether the components were fixated at the time between 500 ms before and 5 s after each component was verbally evoked (Boucheix, Lowe, Putri, & Groff, 2013). The time segmentation was such that the narration regarding that visual location lasted around 5 s. We applied this time-locked analysis to each of the eye-tracking measures used in Experiment 1. Table 6 shows the mean score (and *SD*) for each of the four groups on fixation time, fixation count, average fixation, first fixation duration, and glances count. As with the learning outcome scores, we conducted two-way ANCOVAs for each eye-tracking measure.

With respect to fixation time, the main effect of agent did not reach significance, $F(1, 104) = 3.17, p = .078, \eta_p^2 = .03$ (PA groups: $M = 21.95, SD = 11.12$; No PA groups: $M = 18.76, SD = 9.60$). The main effect of gesture was significant, $F(1, 104) = 94.38, p < .001, \eta_p^2 = .48$, with gesture groups ($M = 27.60, SD = 9.54$) having longer fixation time than no gesture groups ($M = 13.48, SD = 5.49$). The interaction between these two factors was not significant, $F(1, 104) = 1.63, p > .05$.

With respect to fixation count, the main effect of agent was not significant ($F < 1$). The main effect of gesture was significant, $F(1, 104) = 79.10, p < .001, \eta_p^2 = .43$, with gesture groups ($M = 65.92, SD = 18.15$) having a higher fixation count than no gesture groups ($M = 38.39, SD = 13.48$). The interaction between these two factors was not significant ($F < 1$).

With respect to average fixation, the main effect of agent was significant, $F(1, 104) = 3.97, p < .05, \eta_p^2 = .04$, with PA groups ($M = 401, SD = 317$) having longer average fixation time than no PA groups ($M = 307, SD = 181$). The main effect of gesture was

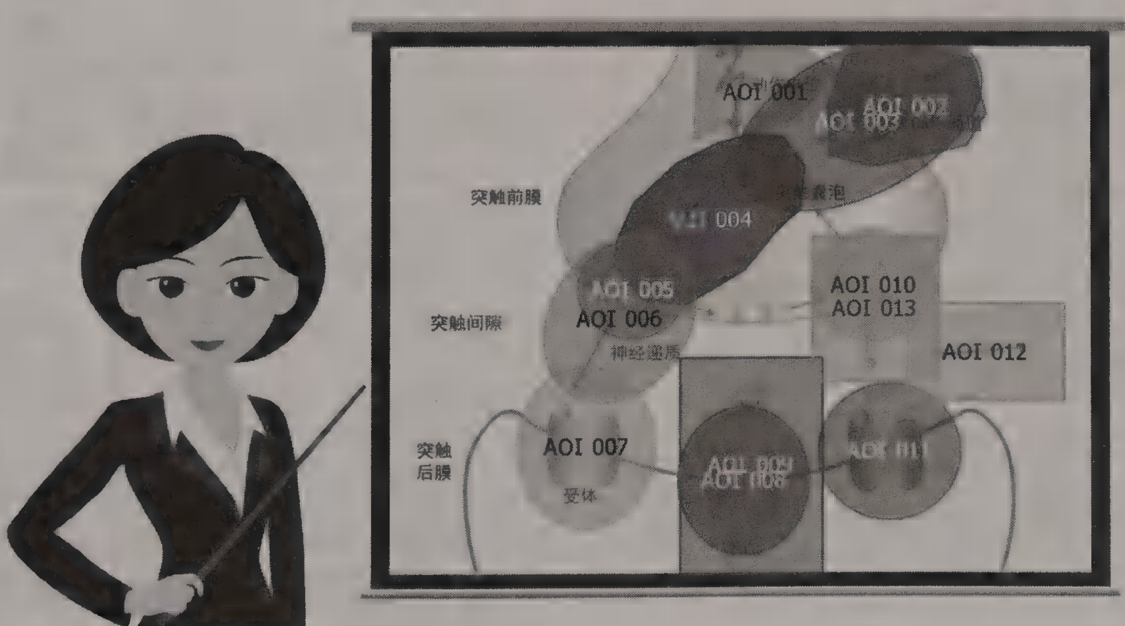

Figure 6. Thirteen temporary areas of interest (AOIs) in Experiment 2. See the online article for the color version of this figure.

Table 6
Mean Scores and SDs on Eye-Tracking Measures for Two Groups in Experiment 2

	PA				No PA			
	Gesture		No gesture		Gesture		No gesture	
Dependent variables	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fixation time (ms)	29,874	9,266	14,033	6,010	25,232	9,415	12,966	5,012
Fixation count	66.37	19.58	38.41	14.61	65.46	16.90	38.38	12.61
Average fixation (ms)	538	391	265	111	422	205	205	56
First fixation duration (ms)	497	355	256	111	379	135	206	58
Glances count	34.81	9.68	22.59	7.77	36.23	8.71	21.83	7.09

Note. PA = pedagogical agent.

significant, $F(1, 104) = 31.30$, $p < .001$, $\eta_p^2 = .23$, with gesture groups ($M = 481$, $SD = 317$) having longer average fixation time than the no gesture groups ($M = 234$, $SD = 91$). The interaction between these two factors was not significant ($F < 1$).

With respect to first fixation duration, the main effect of agent was significant, $F(1, 104) = 4.86$, $p < .05$, $\eta_p^2 = .05$, with agent groups ($M = 376$, $SD = 287$) having longer fixation time than no agent groups ($M = 288$, $SD = 133$). The main effect of gesture was significant, $F(1, 104) = 29.34$, $p < .001$, $\eta_p^2 = .22$, with gesture groups ($M = 439$, $SD = 274$) having longer fixation time than the no gesture groups ($M = 230$, $SD = 90$). The interaction between these two factors was not significant ($F < 1$).

With respect to glances count, the main effect of agent was not significant ($F < 1$). The main effect of gesture was significant, $F(1, 104) = 68.68$, $p < .001$, $\eta_p^2 = .40$, with gesture groups ($M = 35.51$, $SD = 9.15$) having more glances than the no gesture groups ($M = 22.20$, $SD = 7.37$). The interaction between these two factors was not significant ($F < 1$).

Overall, as predicted by the signaling hypothesis, the above results reveal that pointing to areas in the illustration with a pointer effectively guided learners' visual attention to the signaled learning content. Thus, these findings provide strong evidence that having a pointer signal where to look on the illustrations (i.e., in the PA-gesture and No PA-gesture groups) increases the chances that the learner will look at the relevant portion of the illustration as compared with having no pointer (i.e., PA-no-gesture and No PA-no-gesture groups). These results support the proposal that the agent's gesture also plays the role of a signal in guiding learner's visual attention in multimedia learning (Craig et al., 2015; Johnson et al., 2015).

This finding validates the effectiveness of pointing gestures in initiating the first step in multimedia learning—attending to the relevant aspects of the illustration. However, in the next step, to build meaningful learning outcome, learners must exert effort to mentally organize and integrate information. Even though both the PA-gesture group and the No PA-gesture group directed their attention to relevant portions of the illustration better than the control group (No PA-no-gesture), the learning outcome data show that only the PA-gesture group continued on to exert the additional effort needed to build stronger learner outcomes as compared with the control group. This shows evidence for the power of a PA as a social cue that causes learner to process the material more deeply, yielding benefits beyond signaling where to look on the screen.

Experiment 3

Experiments 1 and 2 showed that adding a gesturing PA (but not a motionless PA) to an onscreen lesson effectively guided learner's attention to the relevant elements in the illustration and resulted in improved learning outcomes. Given that creating a highly embodied animated PA takes a lot of time, labor, and cost, Experiment 3 explored whether we can replace an embodied animated PA (i.e., a PA exhibiting eye gaze, body movement, and pointing to key components on the screen as they are being described) with simple physical cues that signal where to look on the screen (i.e., color coding in which key components change color as they are being described). For ease of communication, we refer to the social cues exhibited by the embodied PA as *gesturing*, and we refer to the physical cues of color coding of key components on the screen as *signaling*.

Signaling (or cueing) refers to highlighting the essential parts of the verbal and/or pictorial material in a multimedia, such as using bolding or color to highlight printed text, vocal stress to highlight spoken text, or arrows or color to highlight parts of a graphic (Mayer & Fiorella, 2014; van Gog, 2014). A recent review shows that visual cueing in the form of arrows or coloring or spotlighting can improve student learning of scientific material (van Gog, 2014), although in some cases visual cueing is not as effective as verbal cueing (Mautone & Mayer, 2001). Specifically, some studies without PAs have shown that visual cueing or signaling (such as highlighting key components with arrows, coloring, or spotlighting) could guide the learners' attention in the multimedia environment and in some cases, improve learning outcomes of scientific and technical content (Boucheix & Lowe, 2010; de Koning et al., 2007; Mayer & Fiorella, 2014; van Gog, 2014; Wang et al., 2013).

Method

Participants and design. There were 96 undergraduates who were recruited from a university in central China. Their mean age was 20.0 years ($SD = 2.2$) and 84 of them were women. This experiment used a 2×2 between-subjects design with gesturing (having a gesturing PA that points to key components on the screen with a pointer as they are being described or a nongesturing PA that stands motionless) and signaling (having key components change color as they are being described or not) as factors. The participants were randomly assigned to four groups: 25 in the gesturing and signaling group (gesturing/

signaling), 20 in the gesturing without signaling group (gesturing/no-signaling), 26 in the signaling without gesturing group (no-gesturing/signaling), and 25 in the no gesturing and no signaling group (no gesturing/no-signaling). Example lesson frames for each group are shown in Figure 7. All participants had normal or corrected-to-normal vision and Chinese was their native language. There was no significant difference among the four groups on mean pretest score, $F(3, 92) = 0.31, p > .05$; mean age, $F(3, 92) = 2.40, p > .05$; and proportion of men and women, $\chi^2(3) = 1.48, p > .05$. Participants were majoring in Education (25), Psychology (22), English (17), Chinese (9), Mathematics (8), Computer (4), Politics (3), Physical Education (2), Journalism (2), History (2), Geography (1), and Finance (1).

Materials and apparatus. The materials were same as Experiment 1 except there were four multimedia lessons that described the process of chemical synaptic transmission, either with or without the social cue of a PA pointing to relevant elements in the illustration (i.e., gesturing) and with or without the physical cues of red color guiding students' attention to relevant elements in the illustration (i.e., signaling). As shown in Figure 7, the gesturing/no-signaling version was the same as the PA version in Experiment 1. The no-gesturing/no-signaling version was the same as the PA-no-gesture version in Experiment 2. The gesturing/signaling version had both the social cue of a gesturing PA who pointed to key elements being described in the narration (that we call gesturing) and the physical cue of key components turning red when they were being described in the narration (that we call signaling). The no-gesturing/signaling version had the only physical cue of signaling to guide students' attention.

The pretest questionnaire, retention test, transfer test, and matching test were the same as in Experiment 1. Interrater reliability was $r = .96$ ($p < .001$) for the retention test and $r = .98$ ($p < .001$) for the transfer test. The Cronbach's α for these three tests was 0.76. The apparatus was the same as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except that the participants were randomly assigned to four groups.

Results and Discussion

Posttest scores: How does gesturing and signaling affect learning outcomes? Table 7 shows the means (and *SD*s) of the four groups on the retention test, transfer test, and matching test. To explore the effects of the social cue of gesturing and the physical cue of signaling, and test Hypothesis 3, we performed two-way ANCOVAs with gesturing and signaling as factors and pretest score as a covariate. The same pattern of significant differences was obtained when we conducted ANOVAs.

For the retention test, results showed that the main effect of gesturing was significant, $F(1, 91) = 11.98, p < .01, \eta_p^2 = .12$, in which participants in the gesturing groups ($M = 7.48, SD = 2.58$) outperformed those in the no gesturing groups ($M = 5.75, SD = 2.65$). The main effect of signaling was not significant ($F < 1$). The interaction between gesturing and signaling was significant, $F(1, 91) = 4.03, p < .05, \eta_p^2 = .04$, in which the gesturing group significantly outperformed the no gesturing group when there was no signaling, $F(1, 92) = 11.17, p < .01$, but not when there was signaling, $F(1, 92) = 1.75, p > .05$. Also, signaling did not significantly improve learning whether there was gesturing ($F < 1$) or not, $F(1, 92) = 2.76, p > .05$.

For the transfer test, the main effect of gesturing was significant, $F(1, 91) = 10.70, p < .01, \eta_p^2 = .11$, in which participants in the

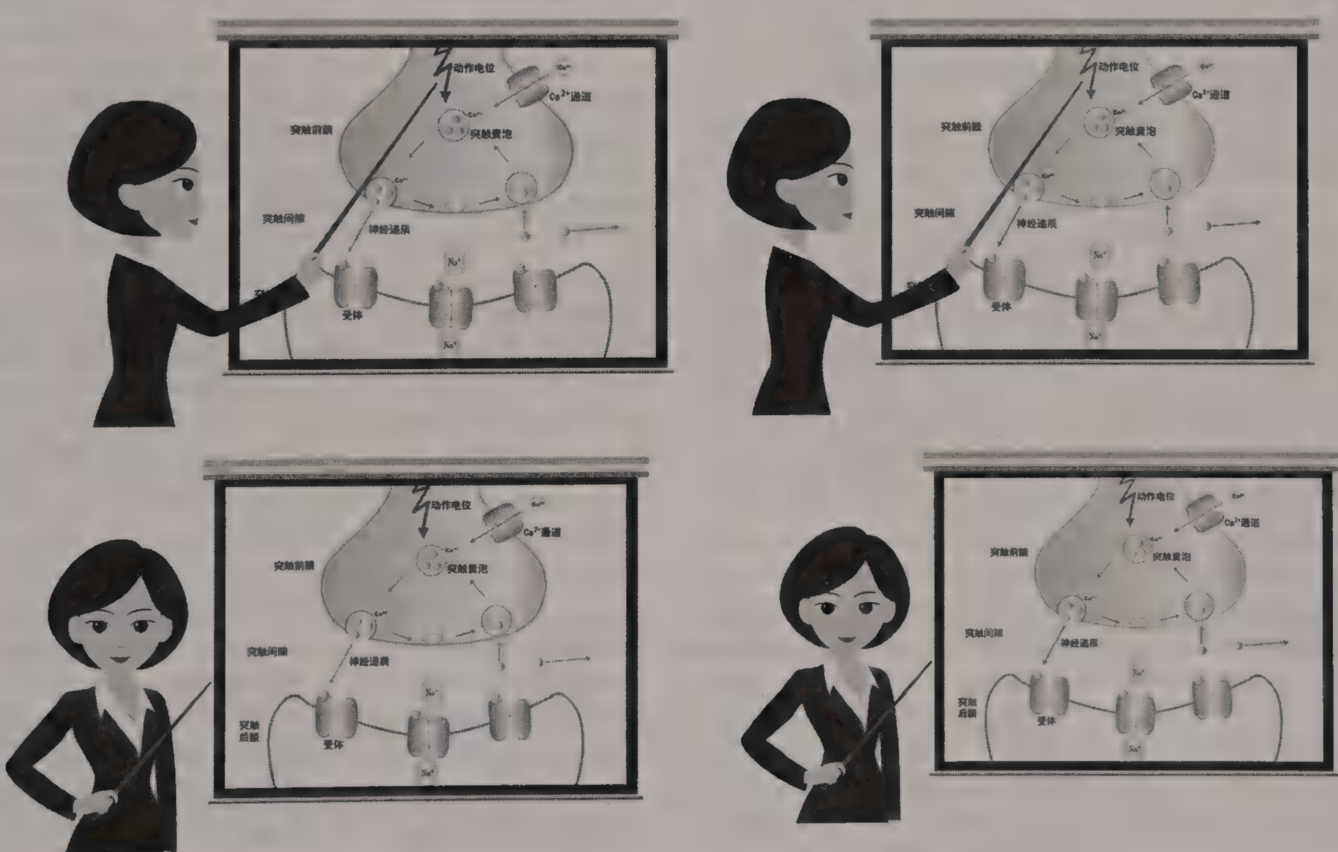

Figure 7. Example frame from animations in Experiment 3. Upper left is gesturing/signaling version, upper right is gesturing/no-signaling version, lower left is no-gesturing/signaling, and lower right is no-gesturing/no-signaling. See the online article for the color version of this figure.

Table 7
Mean Scores on Learning Posttests and SDs for Two Groups in Experiment 3

Dependent variables	Gesturing				No gesturing			
	Signaling		No signaling		Signaling		No signaling	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Retention test	7.33	2.53	7.66	2.71	6.36	2.43	5.11	2.78
Transfer test	4.69	1.96	5.30	2.28	4.58	1.48	3.16	1.91
Matching test	4.96	2.04	5.71	1.23	4.58	1.73	4.34	1.91

gesturing groups ($M = 4.96$, $SD = 2.11$) outperformed those in the no gesturing groups ($M = 3.88$, $SD = 1.83$). The main effect of signaling was not significant, $F(1, 91) = 1.73$, $p > .05$. The interaction between two factors was significant, $F(1, 91) = 13.89$, $p < .001$, $\eta_p^2 = .13$, in which the gesturing group significantly outperformed the no gesturing group when there was no signaling, $F(1, 92) = 14.78$, $p < .001$, but not when there was signaling ($F < 1$). Also, signaling significantly improved learning when there was no gesturing, $F(1, 92) = 6.86$, $p < .05$, but not when there was ($F < 1$).

For the matching test, the main effect of gesturing was significant, $F(1, 91) = 7.63, p < .01, \eta_p^2 = .08$, with the gesturing groups ($M = 5.29$, $SD = 1.75$) outperforming the no gesturing groups ($M = 4.46$, $SD = 1.81$). The main effect of signaling was not significant ($F < 1$). The interaction between the two factors was significant, $F(1, 91) = 5.48, p < .05, \eta_p^2 = .06$, with the gesturing group significantly outperforming the no gesture group when there was no signaling, $F(1, 92) = 6.36, p < .05$, but not when there was signaling ($F < 1$). In addition, signaling did not significantly improve learning whether there was gesturing, $F(1, 92) = 1.48$, $p > .05$, or not ($F < 1$).

Overall, as predicted by the social cue hypothesis, results indicated that adding gesturing to a PA caused students to learn better, particularly when there was no signaling in the form of color coding. This result is similar to previous studies (Craig et al., 2015; Lusk & Atkinson, 2007; Mayer & DaPra, 2012) in which learners performed better when PA exhibited social cues such as gesturing and expression, as compared with a static PA, indicating that the social cue of gesturing has an important influence on learning outcomes in multimedia learning as well as in the real classroom. In contrast, adding the physical cue of color coding as a way of signaling where to look did not produce an overall significant effect, but signaling did improve learning (only on transfer test) when we focus on groups that did not receive gesturing agents. These results show that social cue of gesturing has a stronger impact on learning than signaling in the form of color coding, which is not consistent with signaling hypothesis as articulated in Hypothesis 3. Similar patterns have been reported by other researchers (Johnson et al., 2013, 2015; Moreno et al., 2010). A possible explanation is that the social cue of a gesturing PA not only serves the role of guiding attention, but also primes learners' social stance and facilitates deep processing. In addition, adding signaling in the form of color cueing did not increase learning when there already was a gesturing PA. One reason is that the key elements were already being highlighted by the PA's pointing. Overall, these results show that adding a gesturing PA has positive effects on learning that go beyond simply signaling where to look.

Eye-tracking measures: Does gesturing and signaling affect visual attention during learning? As in Experiment 2, we defined 13 components as AOIs and then a time-locked analysis was conducted on the eye-tracking data. Table 8 shows the mean score (and *SD*) for each of the four groups on fixation time, fixation count, average fixation, first fixation duration, and glances count. As with learning outcome scores, we conducted two-way ANCOVAs for each eye-tracking measure.

With respect to fixation time, the main effect of gesturing was significant, $F(1, 91) = 4.78$, $p < .05$, $\eta_p^2 = .05$, with gesturing groups ($M = 27.16$, $SD = 6.85$) having longer fixation time than no gesturing groups ($M = 24.09$, $SD = 12.54$). The main effect of signaling was significant, $F(1, 91) = 96.63$, $p < .001$, $\eta_p^2 = .52$, with signaling groups ($M = 32.01$, $SD = 7.40$) having longer fixation time than no signaling groups ($M = 18.19$, $SD = 8.05$). The interaction between these two factors was significant, $F(1, 91) = 36.17, p < .001, \eta_p^2 = .28$, with the gesturing group having shorter fixation time than no gesturing group when there was signaling, $F(1, 92) = 8.05, p < .01$, but the gesturing group having longer fixation time than no gesturing group when there was no signaling, $F(1, 92) = 40.74$, $p < .001$. Similarly, the signaling group had longer fixation time than the no signaling group when there was gesturing, $F(1, 92) = 8.09, p < .01$, and when there was no gesturing, $F(1, 92) = 131.83$, $p < .001$.

With respect to fixation count, the main effect of gesturing was significant, $F(1, 91) = 5.88$, $p < .05$, $\eta_p^2 = .06$, with gesturing groups ($M = 68.00, SD = 17.87$) producing a higher fixation count than no gesturing groups ($M = 59.37$, $SD = 26.24$). The main effect of signaling was significant, $F(1, 91) = 50.48$, $p < .001$, $\eta_p^2 = .36$, with signaling groups ($M = 75.65$, $SD = 17.34$) having

Table 8
Mean Scores and SDs on Eye-Tracking Measures for Two Groups in Experiment 3

Dependent variables	Gesturing				No gesturing			
	Signaling		No signaling		Signaling		No signaling	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Fixation time (ms)	29,478	6,153	24,263	6,708	34,442	7,782	13,327	5,277
Fixation count	71.76	16.67	63.30	18.62	79.38	17.46	38.56	15.23
Average fixation (ms)	461	198	411	185	560	280	273	177
First fixation duration (ms)	489	204	391	207	516	279	262	172
Glances count	37.76	7.88	35.30	9.91	38.88	7.43	23.00	7.22

more fixations on relevant AOIs than the no signaling groups ($M = 49.56$, $SD = 20.76$). The interaction between these two factors was significant, $F(1, 91) = 22.50$, $p < .001$, $\eta_p^2 = .20$, in which the gesturing group had a lower fixation count than the no gesturing group when there was signaling, $F(1, 92) = 2.89$, $p = .092$, but the gesturing group had a higher fixation count than the no gesturing group when there was no signaling, $F(1, 92) = 29.78$, $p < .001$. Similarly, the signaling group had a higher fixation count than the no signaling group when there was gesturing, $F(1, 92) = 3.49$, $p = .065$, and when there was no gesturing, $F(1, 92) = 73.33$, $p < .001$.

With respect to average fixation time, the main effect of gesturing was not significant, $F < 1$ ($M_G = 439$, $SD_G = 192$; $M_{NG} = 419$, $SD_{NG} = 287$). The main effect of signaling was significant, $F(1, 91) = 14.29$, $p < .001$, $\eta_p^2 = .14$, with signaling groups ($M = 512$, $SD = 246$) having longer average fixation times than no signaling groups ($M = 334$, $SD = 191$). The interaction between these two factors was significant, $F(1, 91) = 6.98$, $p < .05$, $\eta_p^2 = .07$, reflecting a pattern in which the gesturing group had longer average fixation time than the no gesturing group when there was no signaling, $F(1, 92) = 5.94$, $p < .05$, but not when there was signaling, $F(1, 92) = 2.80$, $p > .05$. Similarly, the signaling group had longer average fixation time than the no signaling group when there was no gesturing, $F(1, 92) = 22.33$, $p < .001$, but not when there was gesturing ($F < 1$).

With respect to first fixation duration, the main effect of gesturing was not significant, $F(1, 91) = 1.22$, $p > .05$ ($M_G = 445$, $SD_G = 209$; $M_{NG} = 391$, $SD_{NG} = 264$). The main effect of signaling was significant, $F(1, 91) = 15.09$, $p < .001$, $\eta_p^2 = .14$, with signaling groups ($M = 503$, $SD = 243$) having longer first fixation duration than no signaling groups ($M = 319$, $SD = 197$). The interaction between these two factors was not significant, $F(1, 91) = 3.03$, $p > .05$.

With respect to glances count, the main effect of gesturing was significant, $F(1, 91) = 11.32$, $p < .01$, $\eta_p^2 = .11$, with the gesturing groups ($M = 36.67$, $SD = 8.82$) engaging in more glances than no gesturing groups ($M = 31.10$, $SD = 10.81$). The main effect of signaling was significant, $F(1, 91) = 30.37$, $p < .001$, $\eta_p^2 = .25$, with signaling groups ($M = 38.33$, $SD = 7.59$) engaging in more glances count than no signaling groups ($M = 28.47$, $SD = 10.44$). The interaction between these two factors was significant, $F(1, 91) = 15.99$, $p < .001$, $\eta_p^2 = .15$, reflecting a pattern in which the gesturing group produced more glances than the no gesturing group when there was no signaling, $F(1, 92) = 30.83$, $p < .001$, but not when there was signaling ($F < 1$). Similarly, the signaling group had a higher glances count than the no signaling group when there was no gesturing, $F(1, 92) = 48.82$, $p < .001$, but not when there was gesturing, $F(1, 92) = 1.68$, $p > .05$.

Overall, based on eye-tracking measures, both gesturing (i.e., having an agent point to relevant parts of the graphic) and signaling (i.e., having relevant parts of the graphic turn red) tended to be helpful for guiding learners' attention, with the poorest performance from the group that received neither type of cue. These results are partially consistent with Experiment 2 in which gesture can guide learner's attention and also are consistent with previous research (Boucheix & Lowe, 2010; de Koning et al., 2007; Ozcelik et al., 2009) where physical cues in multimedia learning can guide learner's attention.

Although both gesturing and signaling were effective in encouraging learners to look at the appropriate part of the screen, previous analyses show that social cues were more effective in encouraging learners to engage in deeper cognitive processing necessary for improved learning outcomes (on retention, transfer, and matching tests). As in Experiment 1, we conclude that attending is only the first step needed for deep learning. For example, Gregory and Hodgson (2012) used the antisaccade task to compare the effects of social cues and physical cues; and they found that social cues automatically activated the oculomotor system, but physical cues did not. Alternatively, social cues may have the evolutionary advantage because they had important effects on the survival and development of human and animals (Langton & Bruce, 1999) rather than physical cues. Given that attending to relevant material is just a useful first step, engaging in deeper processing is vital on learning outcome, particularly transfer tests. Thus, taken together, social cues (i.e., gesturing by an onscreen agent) cannot be replaced completely by physical cues (i.e., signaling by color coding).

General Discussion

Empirical Contributions

The main empirical findings concern the PA hypothesis, the embodiment hypothesis, the image hypothesis, and the signaling hypothesis. Concerning the PA hypothesis (as articulated in Hypotheses 1 and 2), in Experiments 1 and 2, students who received online lessons with a highly embodied PA outperformed students who received the same lesson without a PA on posttest scores of learning outcome and spent more time attending to target material based on eye-tracking measures. In Experiment 2, the benefits of including a PA were found when the PA used human-like gesture including pointing (i.e., high embodiment) but not when it stood motionless (i.e., low embodiment). This finding contributes to the small but growing research base supporting the educational benefits of embodied PAs.

Concerning the embodiment hypothesis (as articulated in Hypothesis 3), in Experiments 2 and 3, there was an embodiment effect in which students who learned with a high embodiment PA outperformed students who learned with low embodiment PA on learning scores and spent more time attending to target material based on eye-tracking measures. Evidence in support of the embodiment effect, including eye-tracking measures, is the main contribution of this set of experiments.

Concerning the image hypothesis, in Experiment 2, there was no image effect because adding the image of a motionless PA (i.e., low embodiment) to an online lesson did not improve performance on learning posttest scores or eye-tracking measures as compared with having no agent. The lack of strong effect for adding a static image of the PA is consistent with previous studies related to the image principle (Mayer, 2014).

Concerning a signaling effect, adding signals (or cues) such as color coding (Experiment 3) or a pointer (Experiment 2) helped direct learner attention during learning based on eye-tracking measures, showing that physical signals were effective in the initial stage of learning. However, this signaling effect for visual attention was not as strong or consistent on measures of learning outcome as the embodiment effect. In Experiment 3, adding the

physical cue of color coding provided modest boosts in some learning outcome measures for students who did not also see a PA (consistent with research on color coding summarized by van Gog, 2014), but not for students who saw a PA who gestured using a pointer, consistent with Hypothesis 4.

Practical Implications

This set of studies supports the embodiment principle: People learn better from an onscreen multimedia lesson when a gesturing PA is added. An important practical implication is that instructional designers should consider adding an onscreen PA who points to key elements in a graphic as they are being described in the narration.

This set of studies does not support the image hypothesis and is consistent with the statement: People do not learn better from an onscreen multimedia lesson when a static image of a pedagogical image is added to the screen. In short, a complementary practical implication is that simply adding the image of a PA is not helpful.

Finally, this study shows that adding additional physical signaling such as color coding of elements in a graphic does not add any benefit when there already is a gesturing PA on the screen. Thus, another complementary practical implication is to not add additional signaling (or cueing) to a gesturing PA.

Theoretical Implications

Overall, the results in this study are consistent with the predictions of social agency theory, which posits that adding social cues such as a gesturing PA to the screen will prime learners to attend to and process the learning content more actively and, therefore, perform better on retention and transfer tests. In the present study, there is evidence that adding a gesturing PA to an onscreen multimedia lesson can prime a social stance in the learner, which engages students in deeper cognitive processing during learning and, therefore, yields meaningful learning outcomes (Mayer et al., 2003). The results of Experiment 3 show that positive effects of adding a gesturing PA go beyond simply providing visual cueing or signaling for where to look on the screen.

Consistent with social agency theory, students who learn with gesturing PAs show more processing of the relevant material during learning—as indicated by more and longer eye fixations. Also consistent with social agency theory, students who learn with gesturing PAs show better learning outcomes, presumably because they have processed the incoming material more deeply. These findings show there is a social side to online learning that can be tapped through appropriate instructional design to improve student learning. The eye-tracking data (e.g., increased fixation time for the PA group) are consistent with the idea the social cues exhibited by the gesturing PA cause the learner to work harder during learning leading to better learning outcomes (e.g., better posttest scores).

On the other hand, these results are not consistent with the idea that gesturing PAs create extraneous cognitive load that detracts from learning. It should be noted that adding a PA to the screen did attract a small amount of the learners' attention during learning, especially at the stage of introduction to the topic, but this was not enough to harm learning.

Methodological Implications

According to Mayer and DaPra (2012), research on PAs is faced with three challenging measurement questions: (a) measuring the learner's social stance, (b) measuring the learner's cognitive processing during learning, (c) and measuring the learner's meaningful learning outcome. In this study, we used subjective self-ratings to measure the learner's social stance, but it should be noted that this technique is problematic because it is not objective and proved to be insensitive to differences in social perception. Thus, future studies should use more reliable physiological indicators to objectively reflect the learner's social stance, such as by measuring brain activity during learning with minimal intrusion with functional near-infrared spectroscopy (fNIRS) and electroencephalograph (EEG).

In this study we used eye-tracking to measure the learner's cognitive processing during learning, particularly, the learner's distribution of visual attention during learning. We found eye-tracking methodology to be a useful way to measure the learner's attention during multimedia learning.

Finally, we measured learning outcomes with a multileveled battery consisting of retention and matching tests to measure memorization of the presented material and a transfer test to measure understanding. In a recent meta-analysis (Wang et al., 2017) found that PAs had different effects on different measurements, so it is fruitful to use a multileveled battery as in this study.

Limitations and Future Directions

This research also had some limitations that further research should address. First, this study focused on college students as subjects, but there are great differences in the physical and mental development of college students versus younger students. In a meta-analysis, Schroeder et al. (2013) showed that the learner's age is an important variable that influences learning effects. Thus, future research should include K-12 school students as subjects to explore the effect of social cues on learning across a greater age span.

Second, in this study, all three experiments used the same learning material concerning the chemical synaptic transmission process, which belongs to the field of science. Meta-analysis have found that the content of the learning material is an important moderating factor that influenced the effect of PAs (Schroeder et al., 2013; Wang et al., 2017), in which PAs had a positive effect on learning for scientific fields (e.g., mathematics, chemistry, biology, and physics), but not in the humanities. Thus, future research should use materials from different disciplines to examine the generalizability of the effects of social cues.

Third, learner's prior knowledge has been shown to be an important cognitive feature that affects learning and performance (Kalyuga, 2007, 2014). An expertise reversal effect was found in multiple domains of multimedia learning, in which instructional techniques that are highly effective with inexperienced learners can lose their effectiveness and even have negative consequences when used with more experienced learners (Kalyuga, 2007, 2014; Kalyuga, Ayres, Chandler, & Sweller, 2003). An important issue for future research is how different experience levels of learners would social cue have different effects. Because this study did not aim to explore the role of the learner's prior knowledge, this study used the learner's prior knowledge as a covariate. Future research

can simultaneously use high experienced learners and low experienced learners as subjects to explore the impact of prior knowledge on the effectiveness of social cues.

Fourth, the studies reported in this paper, like most of the studies with PAs, involve a short intervention, so a worthwhile direction for future research is to examine whether the effects persist in longer and more complex learning situations.

Fifth, a useful future direction is to design studies that allow for a more fine-grained analysis of how the learning of specific content is linked to specific cues during instruction. In particular, we would want to know whether having a gesturing PA helps students attend to and learn particular aspects of the instructional content that are directly linked to the PA's pointing gestures. The dependent measures in the present study were not designed to address this issue.

Finally, this study only uses eye tracking to explore the cognitive processing during learning with an agent. This technique does not examine brain activity when students learn with and without a PA. Future research could use techniques such as fNIRS, fMRI, and EEG to explore the neural activity of the brain during learning.

References

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology, 94,* 416–427. http://dx.doi.org/10.1037//0022-0663.94.2.416

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology, 30,* 117–139. http://dx.doi.org/10.1016/j.cedpsych.2004.07.001

Bailenson, J. N., Swinth, K., Hoyt, C., & Persky, S. (2005). The independent and interactive effects of embodied agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence, 14,* 379–393. http://dx.doi.org/10.1162/105474605774785235

Baylor, A., & Kim, S. (2009). Designing nonverbal communication for pedagogical agents: When less is more. *Computers in Human Behavior, 25,* 450–457. http://dx.doi.org/10.1016/j.chb.2008.10.008

Baylor, A. L., & Ryu, J. (2003). The effect of image and animation in enhancing pedagogical agent persona. *Journal of Educational Computing Research, 28,* 373–394. http://dx.doi.org/10.2190/V0WQ-NWGN-JB54-FAT4

Boucheix, J. M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction, 20,* 123–135. http://dx.doi.org/10.1016/j.learninstruc.2009.02.015

Boucheix, J. M., Lowe, R. K., Putri, D. K., & Groff, J. (2013). Cueing animations: Dynamic signaling aids information extraction and comprehension. *Learning and Instruction, 25,* 71–84. http://dx.doi.org/10.1016/j.learninstruc.2012.11.005

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents.* Cambridge, MA: MIT Press.

Choi, S., & Clark, R. E. (2006). Cognitive and affective benefits of an animated pedagogical agent for learning English as a second language. *Journal of Educational Computing Research, 34,* 441–466. http://dx.doi.org/10.2190/A064-U776-4208-N145

Clark, R. E., & Choi, S. (2005). Five design principles for experiments on the effects of animated pedagogical agents. *Journal of Educational Computing Research, 32,* 209–225. http://dx.doi.org/10.2190/7LRM-3BR2-44GW-9QQY

Craig, S. D., Twyford, J., Irigoyen, N., & Zipp, S. A. (2015). A test of spatial contiguity for virtual human's gestures in multimedia learning environments. *Journal of Educational Computing Research, 53,* 3–14. http://dx.doi.org/10.1177/0735633115585927

Dehn, D. M., & Van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies, 52,* 1–22. http://dx.doi.org/10.1006/ijhc.1999.0325

de Koning, B. B., & Tabbers, H. K. (2013). Gestures in instructional animations: A helping hand to understanding non-human movements? *Applied Cognitive Psychology, 27,* 683–689.

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology, 21,* 731–746. http://dx.doi.org/10.1002/acp.1346

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010a). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction, 20,* 111–122. http://dx.doi.org/10.1016/j.learninstruc.2009.02.010

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010b). Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education, 55,* 681–691. http://dx.doi.org/10.1016/j.compedu.2010.02.027

Dirkin, K. H., Mishra, P., & Altermatt, E. (2005). All or nothing: Levels of sociability of a pedagogical software agent and its impact on student perceptions and learning. *Journal of Educational Multimedia and Hypermedia, 14,* 113–127.

Dunsworth, Q., & Atkinson, R. K. (2007). Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers & Education, 49,* 677–690. http://dx.doi.org/10.1016/j.compedu.2005.11.010

Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology, 108,* 528–546. http://dx.doi.org/10.1037/edu0000065

Frechette, C., & Moreno, R. (2010). The roles of animated pedagogical agents' presence and nonverbal communication in multimedia learning environments. *Journal of Media Psychology: Theories, Methods, and Applications, 22,* 61–72. http://dx.doi.org/10.1027/1864-1105/a000009

Gregory, N. J., & Hodgson, T. L. (2012). Giving subjects the eye and showing them the finger: Socio-biological cues and saccade generation in the anti-saccade task. *Perception, 41,* 131–147. http://dx.doi.org/10.1068/p7085

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), Vol. 3, pp. 41–58). *Syntax and semantics* New York, NY: Academic Press.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90,* 414–434. http://dx.doi.org/10.1037/0022-0663.90.3.414

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review, 6,* 27–54. http://dx.doi.org/10.1016/j.edurev.2010.07.004

Holmes, J. (2007). Designing agents to support learning by explaining. *Computers & Education, 48,* 523–547. http://dx.doi.org/10.1016/j.compedu.2005.02.007

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20,* 172–176. http://dx.doi.org/10.1016/j.learninstruc.2009.02.013

Jin, S. A. A. (2010). The effects of incorporating a virtual agent in a computer-aided test designed for stress management education: The mediating role of enjoyment. *Computers in Human Behavior, 26,* 443–451. http://dx.doi.org/10.1016/j.chb.2009.12.003

Johnson, A. M., Ozogul, G., Moreno, R., & Reisslein, M. (2013). Pedagogical agent signaling of multiple visual engineering representations:

The case of the young female agent. *Journal of Engineering Education, 102,* 319–337. http://dx.doi.org/10.1002/jee.20009

Johnson, A. M., Ozogul, G., & Reisslein, M. (2015). Supporting multimedia learning with visual signalling and animated pedagogical agent: Moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31,* 97–115. http://dx.doi.org/10.1111/jcal.12078

Johnson, W. L., & Lester, J. C. (2016). Face-to-face interaction with pedagogical agents, twenty years later. *International Journal of Artificial Intelligence in Education, 26,* 25–36. http://dx.doi.org/10.1007/s40593-015-0065-9

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11,* 47–78. Retrieved from http://webpages.uncc.edu/tbarnes2/ITSspr06/papers/ped-agents.pdf

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19,* 509–539. http://dx.doi.org/10.1007/s10648-007-9054-3

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 576–597). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.028

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38,* 23–31. http://dx.doi.org/10.1207/S15326985EP3801_4

Kim, Y., & Baylor, A. L. (2016). Research-based design of pedagogical agent roles: A review, progress, and recommendations. *International Journal of Artificial Intelligence in Education, 26,* 160–169. http://dx.doi.org/10.1007/s40593-015-0055-y

Langton, S. R., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6,* 541–567. http://dx.doi.org/10.1080/135062899394939

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (pp. 359–366). New York, NY: ACM.

Lester, J. C., Towns, S. G., & Fitzgerald, P. J. (1998). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. *International Journal of Artificial Intelligence in Education, 10,* 278–291. Retrieved from http://lutes.upmc.fr/delozanne/2012-2013/methodo/LesterIJAIED99.pdf

Li, W. J., Tong, Y., Wang, F. X., Kang, S. J., Liu, H. S., & Yang, C. (2016). Effect of animation pedagogical agent in multimedia learning: The role of learner' s experience and agent preference. *Psychological Development and Education, 32,* 453–462. http://dx.doi.org/10.16187/j.cnki.issn1001-4918.2016.04.09

Louwerse, M. M., Graesser, A. C., Mcnamara, D. S., & Lu, S. L. (2009). Embodied conversational agents as conversational partners. *Applied Cognitive Psychology, 23,* 1244–1255. http://dx.doi.org/10.1002/acp.1527

Lusk, M. M., & Atkinson, R. K. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology, 21,* 747–764. http://dx.doi.org/10.1002/acp.1347

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93,* 377–389. http://dx.doi.org/10.1037/0022-0663.93.2.377

Mautone, P. D., & Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology, 99,* 640–652. http://dx.doi.org/10.1037/0022-0663.99.3.640

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511811678

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20,* 167–171. http://dx.doi.org/10.1016/j.learninstruc.2009.02.012

Mayer, R. E. (Ed.). (2014). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment. *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–370). New York, NY: Cambridge University Press.

Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied, 18,* 239–252. http://dx.doi.org/10.1037/a0028616

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96,* 389–395. http://dx.doi.org/10.1037/0022-0663.96.2.389

Mayer, R. E., & Fiorella, L. (2014). Principle for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). Cambridge, United Kingdom: Cambridge University Press.

Mayer, R. E., Heiser, H., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93,* 187–198. http://dx.doi.org/10.1037/0022-0663.93.1.187

Mayer, R. E., Sobko, K., & Mautone, P. D. (2003). Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology, 95,* 419–425. http://dx.doi.org/10.1037/0022-0663.95.2.419

McGregor, K. K. (2008). Gesture supports children's word learning. *International journal of speech-language pathology, 10,* 112–117. http://dx.doi.org/10.1080/17549500801905622

McGregor, K. K., Rohlfing, K. J., Bean, A., & Marschner, E. (2009). Gesture as a support for word learning: The case of under. *Journal of Child Language, 36,* 807–828. http://dx.doi.org/10.1017/S0305000908009173

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011a). Polite web-based intelligent tutors: Can they improve learning in classrooms? *Computers & Education, 56,* 574–584. http://dx.doi.org/10.1016/j.compedu.2010.09.019

McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011b). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies, 69,* 70–79. http://dx.doi.org/10.1016/j.ijhcs.2010.09.001

Moreno, R. (2005). Multimedia learning with animated pedagogical agents. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 507–524). New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511816819.032

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31,* 186–207. http://dx.doi.org/10.1016/j.cedpsych.2005.05.002

Moreno, R., & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 92,* 724–733. http://dx.doi.org/10.1037/0022-0663.92.4.724

Moreno, R., & Mayer, R. E. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96,* 165–173. http://dx.doi.org/10.1037/0022-0663.96.1.165

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents. *Cognition and Instruction, 19,* 177–213. http://dx.doi.org/10.1207/S1532690XCI1902_02

Moreno, R., Reislein, M., & Ozogul, G. (2010). Using virtual peers to guide visual attention during learning. *Journal of Media Psychology:*

Theories, Methods, and Applications, 22, 52–60. http://dx.doi.org/10.1027/1864-1105/a000008

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press.

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education, 53,* 445–453. http://dx.doi.org/10.1016/j.compedu.2009.03.002

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38,* 1–4. http://dx.doi.org/10.1207/S15326985EP3801_1

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. http://dx.doi.org/10.1037//0033-2909.124.3.372

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 62,* 1457–1506. http://dx.doi.org/10.1080/17470210902816461

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places.* New York, NY: Cambridge University Press.

Schroeder, N. L., & Adesope, O. O. (2014). A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education, 46,* 229–251. http://dx.doi.org/10.1080/15391523.2014.888265

Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research, 49,* 1–39. http://dx.doi.org/10.2190/EC.49.1.a

Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science, 16,* 85–89. http://dx.doi.org/10.1111/j.0956-7976.2005.00786.x

Sweller, J., Ayres, P., & Kalyuga, P. (2011). *Cognitive load theory.* New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4419-8126-4

Unal-Colak, F., & Ozan, O. (2012). The effect of animated agents on students' achievement and attitudes. *Journal of Distance Education, 13,* 96–111.

van Gog, T. (2014). The Signaling (or Cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). Cambridge, United Kingdom: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.014

Veletsianos, G., & Russell, G. S. (2014). Pedagogical agents. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 759–769). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4614-3185-5_61

Wang, F. X., Duan, Z. H., & Zhou, Z. K. (2013). Attention guidance in multimedia learning: The role of cueing. *Advances in Psychological Science, 21,* 1430–1440. http://dx.doi.org/10.3724/SP.J.1042.2013.01430

Wang, F. X., Li, W. J., Xie, H. P., & Liu, H. S. (2017). Is pedagogical agent in multimedia learning good for learning? A meta-analysis. *Advances in Psychological Science, 25,* 12–28. http://dx.doi.org/10.3724/SP.J.1042.2017.00012

Yılmaz, R., & Kılıç-Çakmak, E. (2012). Educational interface agents as social models to influence learner achievement, attitude and retention of learning. *Computers & Education, 59,* 828–838. http://dx.doi.org/10.1016/j.compedu.2012.03.020

Yung, H. I., & Paas, F. (2015). Effects of cueing by a pedagogical agent in an instructional animation: A cognitive load approach. *Journal of Educational Technology & Society, 18,* 153–160. Retrieved from http://ifets.info/journals/18_3/12.pdf

(*Appendix follows*)

Appendix

Idea Units for Scoring in Retention Test

Chemical synaptic transmission between neurons mainly occurs in the presynaptic membrane, synaptic gap and the postsynaptic membrane.

The transfer process needs more steps to complete.

Action potentials, the presynaptic neurons are generated, and transmitted to the presynaptic membrane of nerve terminals.

The arrival of action potentials induces depolarization of presynaptic membrane.

Thus, this intensifies the voltage gated Ca^{2+} channel on the presynaptic membrane, and permeability of Ca^{2+} is enhanced.

At this point, Ca^{2+} in the extracellular enters into the presynaptic membrane through the channel, which leads to increasing the concentration of Ca^{2+} in the presynaptic membrane.

The entry of Ca^{2+} may prompt the synaptic vesicle to move to the presynaptic membrane, and synaptic vesicle fuses with presynaptic membrane, then a cleft appears in the presynaptic membrane.

The neurotransmitter in the synaptic vesicle is released into the synaptic gap through the role of the cell.

These neurotransmitters arrive at the postsynaptic membrane by diffusion, and are combined with specific receptors on the postsynaptic membrane.

The combination of neurotransmitters and receptors changes ion's permeability of the postsynaptic membrane, and some ion channels open.

Ions begin to move across the membrane, for example, Na^{+} flows into the postsynaptic membrane, and changes the membrane potential of the postsynaptic membrane, which eventually leads to the postsynaptic potential depolarization or super polarization.

To compensate for the reduction in the number of synaptic vesicles, new vesicles will be re-produced under the action of the related proteins on the presynaptic membrane.

The released neurotransmitter has an inactivation mechanism, it mainly includes three ways:

First is enzyme degradation.

Neurotransmitters that combined with receptor in the synaptic cleft, are rapidly degraded by neurotransmitter enzyme

Second is the diffusion.

That is, a part of neurotransmitters leaves the synapse through passive diffusion.

Third is to reuptake.

That is, another part of neurotransmitters is re-ingested in the presynaptic membrane.

Received December 14, 2016
Revision received May 25, 2017
Accepted June 18, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 269–288

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000214

Daily Autonomy Supporting or Thwarting and Students' Motivation and Engagement in the High School Science Classroom

Erika A. Patall, Rebecca R. Steingut, Ariana C. Vasquez, Scott S. Trimble, Keenan A. Pituch, and Jen L. Freeman
University of Texas at Austin

This diary study provided the first classroom-based empirical test of the relations between student perceptions of high school science teachers' various autonomy supporting and thwarting practices and students' motivation and engagement on a daily basis over the course of an instructional unit. Perceived autonomy supporting practices were hypothesized to predict autonomous motivation and engagement outcomes, while perceived autonomy thwarting practices were hypothesized to predict controlled motivation and disaffection outcomes. In line with this prediction, multilevel modeling results based on regular reports of 208 high school students in 41 science classes across 6 weeks suggested that 4 perceived daily supports (choice provision, consideration for student preferences and interests, rationales for importance, and question opportunities) and 1 daily thwart (use of uninteresting activities) predicted changes in daily autonomous motivation and engagement. In contrast, changes in students' daily controlled motivation and disaffection were predicted primarily by 3 perceived daily thwarts (controlling messages, suppression of student perspectives, and use of uninteresting activities). Results also suggested that practices interacted such that the perception of thwarts generally bolstered desirable daily relationships between perceived supports and students' motivation and the perception of supports generally mitigated undesirable daily relationships between thwarts and motivation. Supplemental exploratory results suggested that the effects of choice and suppression of student perspectives may be heterogeneous and depend on the outcome and/or the presence of other practices. Implications of the findings are discussed.

Educational Impact and Implications Statement

The results of a 6-week classroom-based diary study with 208 high school students in 41 science classes suggested that students' autonomous motivation and engagement increased (since the prior class day) on days when students perceived their teachers to support their autonomy by providing choices, considering their preferences and interests in course activities, communicating rationales for the importance of activities, providing opportunities to ask questions, or avoiding uninteresting activities. In contrast, controlled motivation and disaffection increased on days when students' perceived their teachers to thwart their autonomy by using controlling messages, suppressing student perspectives, or using uninteresting activities. Students' perceptions that teachers' used thwarting practices simultaneously with supportive practices bolstered the desirable relationship between perceived supports and motivation, and mitigated the undesirable relationship between thwarts and motivation. Results suggest the importance of focusing motivation interventions on training high school teachers to implement specific daily practices geared at supporting students' experience of autonomy and minimizing the use of specific thwarting practices to both promote autonomous motivation and engagement and reduce controlled motivation and disaffection. Results highlight the importance of targeting a profile of autonomy-relevant practices that teachers use each day when attempting to maximize student motivation and engagement.

Keywords: autonomy support, autonomy thwart, teaching practice, motivation, engagement

A distressing pattern consistently found in education research is that motivation and engagement decline across grades, with the lowest levels among high school students, and from the start to the end of the school year within secondary classrooms (e.g., Eccles et al., 1993; Harter, 1981; Lepper, Corpus, & Iyengar, 2005; Skinner Furrer, Marchand, & Kindermann, 2008). Moreover, the steepest declines may occur for science, technology, engineering, and mathematics (STEM) fields (e.g., Gottfried, Fleming, & Gottfried, 2001; Gottfried, Marcoulides, Gottfried, & Oliver, 2009), as the percentage of students studying and earning degrees in nearly all STEM fields has remained stable or declined over time (Maltese & Tai, 2011; Organization for Economic Co-operation and Development, 2006). This decline is troubling given extensive evidence that motivation and engagement are central to learning and achievement (e.g., Archambault, Janosz, Fallu, & Pagani, 2009; Hughes, Luo, Kwok, & Loyd, 2008; Lepper et al., 2005; Murayama, Pekrun, Lichtenfeld, & vom Hofe, 2012; Willingham,

This article was published Online First June 5, 2017.

Erika A. Patall, Rebecca R. Steingut, Ariana C. Vasquez, Scott S. Trimble, Keenan A. Pituch, and Jen L. Freeman, Department of Educational Psychology, University of Texas at Austin.

This research was supported by a grant from the William T. Grant Foundation (Project 180042).

Correspondence concerning this article should be addressed to Erika A. Patall, who is now at Rossier School of Education, University of Southern California, 3470 Trousdale Parkway, Waite Phillips Hall, Los Angeles, CA 90089-4036. E-mail: patall@rossier.usc.edu

Pollack, & Lewis, 2002). Moreover, the increasing demand for individuals with knowledge in STEM areas in the current global marketplace (see, e.g., Bureau of Labor Statistics, 2011) make addressing low motivation and engagement in science classrooms particularly important.

Given these circumstances, an important goal of educational and psychological research is to understand how to structure teacher practices and the classroom environment to support students' motivation and engagement and prevent declines especially common at the secondary level and in STEM areas. A substantial body of research grounded in self-determination theory (Ryan & Deci, 2000) has suggested that teachers who are perceived to engage in practices that are supportive of students' experiences of autonomy facilitate optimal functioning in the form of autonomous motivation and engagement (e.g., Assor, Kaplan, & Roth, 2002; Patall, Dent, Oyer, & Wynn, 2013; Reeve & Jang, 2006; Reeve, Jang, Carrell, Jeon, & Barch, 2004). In contrast, controlling teacher practices that thwart students' experiences of autonomy predict controlled motivation, which is driven by external consequences, and maladaptive functioning (e.g., Assor et al., 2002; De Meyer, et al., 2014; Haerens, Aelterman, Vansteenkiste, Soenens, & Van Petegem, 2015; Reeve & Jang, 2006). While autonomy supporting and thwarting is important across contexts, the increasing need for autonomy and independence as students enter adolescence (Eccles et al., 1993; Erikson, 1968) make understanding the effects of teachers' use of autonomy relevant strategies particularly important in the context of secondary school classrooms. Likewise, given that discovery and innovation, recognition of ambiguity, and learning from past discoveries and failure are all central values of science (e.g., Allchin, 1999; Bartos & Lederman, 2014; Kuhn, 1962), support for personal autonomy would seem to be particularly important to science education.

However, limitations in the research on students' experiences of autonomy relevant teaching persist. In particular, the individual practices thought to support or thwart autonomy have been given inadequate attention in research based in a classroom context (and in STEM classes in particular) beyond retrospective, single survey, and cross-sectional designs. Moreover, existing research has failed to investigate within academic classroom contexts the extent to which students' experiences of various individual autonomy supporting and thwarting practices predict distinct motivation and engagement outcomes and the extent to which students' perceptions of autonomy supporting and thwarting interact to affect motivation and engagement. The current study sought to address these gaps by investigating the links between high school science students' perceptions of several autonomy relevant teaching practices and their motivation and engagement in a diary study that made use of repeated daily student reports across a 6-week instructional unit. The two main goals of this investigation were: (a) to examine the relationships between students' perceptions of a set of teaching strategies routinely identified as autonomy supporting or thwarting with students' daily autonomous and controlled motivation, engagement, and disaffection in authentic high school science classrooms and (b) to explore the extent to which perceived supportive and thwarting practices interact to predict students' daily motivation and engagement. The chosen design in which perceived teacher practice and students' motivation and engagement was assessed repeatedly over class days provided an opportunity to collect strong evidence regarding the predictive role of daily perceptions of teacher practice in explaining students' daily motivation and engagement in class.

Teacher Practices That Support or Thwart Autonomy

According to self-determination theory, autonomy, or the experience that one's behavior is volitional and self-endorsed, is central to adaptive functioning and well-being as one of three fundamental human needs, along with needs for competence and relatedness (e.g., Ryan & Deci, 2000). The experience of being controlled is the logical opposite of autonomy, reflecting the perception that behavior is coerced by an external force (e.g., by a teacher's directive or an offer of a reward), is done out of feelings of pressure, obligation, or guilt, or is done because of a lack of choice. Along these lines, research suggests that satisfying the need for autonomy is associated with engagement, well-being, and highly desirable internal forms of motivation (e.g., intrinsic motivation), while experiencing frustration of the need for autonomy is associated with poorer well-being and less desirable forms of motivation that are focused on acquiring rewards or avoiding undesirable consequences (e.g., Bartholomew, Ntoumanis, Ryan, Bosch, & Thøgersen-Ntoumani, 2011; Haerens et al., 2015; Patall et al., 2013; Reeve & Jang, 2006).

Teachers' practices and more proximally, student perceptions of teacher practice, predict students' autonomy satisfaction or frustration, and in turn, the nature of their motivation, engagement, well-being, and achievement (e.g., Assor, Kaplan, Kanat-Maymon, & Roth, 2005; Assor et al., 2002; Haerens et al., 2015; Jang, Kim, & Reeve, 2012; Patall, Cooper, & Wynn, 2010; Patall et al., 2013; Reeve & Jang, 2006; Reeve et al., 2004; Skinner & Belmont, 1993; Soenens, Sirens, Vansteenkiste, Dochy, & Goossens, 2012). Autonomy support in the classroom context reflects a motivational approach in which teachers identify, nurture, and develop students' inner motivational resources so that students perceive themselves as the initiator of their actions (Reeve, 2009). Autonomy supportive teachers are conceptualized as offering choices, encouraging students to work in their own way or at their own pace, and open and responsive to students' opinions and questions. Although such teachers attempt to structure course activities around students' interests whenever possible, they also provide meaningful rationales to explain the usefulness or importance of even "boring" course activities (see Reeve, 2009; Reeve & Jang, 2006; or Su & Reeve, 2011 for a review of autonomy supportive practices).

In contrast, controlling teachers thwart autonomy in that they are perceived to be dismissive of student perspectives and to pressure students to think, act, or feel in particular ways (Reeve, 2009). Relatively fewer studies have addressed practices thought to thwart autonomy or students' perceptions of controlling practices. However, explicitly controlling language (e.g., "you must" or "you should"), commands that pressure students to act in teacher sanctioned ways, rationales that emphasize the external consequences of compliance, suppression of students' questions and opinions, and the assignment of activities that appear meaningless or uninteresting are routinely included among practices expected to thwart students' experiences of autonomy (e.g., Assor et al., 2002, 2005; Reeve, 2009; Reeve & Jang, 2006).

Differential Associations for Autonomy Relevant Teacher Practices

Autonomy support and control have often been conceptualized as being on opposite ends of a continuum. However, more recently, theory and research has increasingly suggested that controlling behaviors cannot by equated with infrequent acts of autonomy support. Rather, research now suggests that students' perceptions of practices that thwart autonomy have a rather modest negative correlation with perceived practices that support autonomy and yield distinct effects (e.g., Bartholomew et al., 2011; Haerens et al., 2015; Jang et al., 2016). While contexts that support autonomy unlock internal motivational resources that allow an individual to thrive, contexts that thwart autonomy can lead to defensive reactions that promote externally focused motivation and ill-being (Deci & Ryan, 2000). Consistent with this, Bartholomew and colleagues (2011) found that perceptions of autonomy-supportive coaching was most closely related to athletes' daily experiences of need satisfaction, and in turn, daily psychological well-being, while perceptions of controlling coaching was most closely related to daily need thwarting, and in turn, daily psychological and physical ill-being. Using a retrospective survey, Assor and colleagues (2002) found that students' perceptions that teachers provided choices and fostered students' understanding of the relevance of course activities primarily predicted emotional, cognitive, and behavioral engagement, while perceptions that teachers intruded on students' behavior and suppressed student perspectives primarily predicted students' negative affect. Also in retrospective, cross-sectional research, Haerens and colleagues (2015) found that perceptions of physical education teachers' autonomy support were primarily related to autonomous motivation via need satisfaction, while perceptions of physical education teachers' controlling practice was primarily related to controlled motivation and amotivation via need frustration. Finally, Jang and colleagues (2016) found that Korean high school students' perceptions that teachers supported their autonomy predicted changes in need satisfaction that predicted changes in engagement over the course of a school semester, while perceived teacher control predicted changes in need frustration and subsequent disengagement.

Implicit in these findings and our discussion is the differentiated view of motivation and engagement outcomes that motivation scholars have come to accept. Self-determination theory differentiates more autonomous and more controlled forms of motivation. Intrinsic and identified forms of motivation represent more autonomous forms in which the regulation of actions is incited by the inherent satisfaction, interest, or enjoyment that a task brings (intrinsic) or one's personal value for tasks (identified). Introjected and extrinsic motivation represent more controlled or external forms of motivation in which action is driven by internally controlling consequences such as feelings of guilt, shame, or pride (introjected) or the desire to obtain rewards and avoid punishment from the environment (extrinsic; e.g., Ryan & Deci, 2000). Moreover, autonomous forms of motivation (intrinsic and identified) are particularly desirable in the classroom because research has routinely indicated that they are linked with a variety of desirable and adaptive outcomes, including creativity, academic engagement, deep conceptual learning strategies, and academic achievement (e.g., Corpus, McClintic-Gilbert, & Hayenga, 2009; Lepper, Corpus, & Iyengar, 2005; Otis, Grouzet, & Pelletier, 2005; Walker, Greene, & Mansell, 2006). In contrast, more extrinsic forms of motivation (introjected and extrinsic) are often linked with negative outcomes, including maladaptive learning strategies and attitudes, anxiety, poorer ability to cope with challenges, poor academic achievement, and even school drop-out (e.g., Lepper et al., 2005; Ryan & Connell, 1989; Vansteenkiste, Zhou, Lens, & Soenens, 2005; Walker et al., 2006).

The contrast between engagement and disaffection represents a similar juxtaposition of more and less desirable functioning. Engagement is typically conceptualized as a motivational construct that has a behavioral dimension that includes effort, persistence, intensity, and perseverance in the face of obstacles, an emotional dimension that includes enthusiasm, enjoyment, fun, and other positive emotions, and a cognitive dimension that includes attention to and regulation of the learning and thinking process (e.g., Skinner, Kindermann, Connell, & Wellborn, 2009). The opposite of engagement is disaffection, disengagement, or helplessness. Disaffection is not merely low levels of engagement of the various types. Rather, it is often operationalized in its behavioral form as giving up, just going through the motions, passivity, and lack of initiation, and in its emotional form as boredom, apathy, frustration, discouragement, or dejection (e.g., Miceli & Castelfranchi, 2000; Skinner et al., 2009).

In line with this current conception of the nature of teachers' motivating style and students' classroom experience, the dual-process model within a self-determination theory framework (e.g., Jang, Kim, & Reeve, 2016) explicitly asserts this differentiated view of teacher practice, student motivation, and student engagement. That is, teachers' motivating practice reflects the distinct processes of both perceived autonomy support and perceived autonomy thwarting. Student motivation and engagement can likewise be differentiated into need satisfaction, autonomous motivation, and engagement on the one hand, and need frustration, controlled motivation, and disaffection on the other hand. Thus, the dual-process model acknowledges that while the autonomy supportive teacher practices are likely to explain students' need satisfaction, autonomous motivation, and engagement, autonomy thwarting teacher practices explain students feelings of being controlled, frustrated, and disengaged.

All things considered, it would seem important to examine the extent to which perceptions of both autonomy supporting and thwarting practices differentially predict motivation and engagement outcomes, as each set of perceived practices are likely to differentially predict students' autonomous motivation and engagement versus controlled motivation and disaffection. However, despite the progress made in research focused on understanding autonomy relevant teaching behaviors, a number of limitations persist. Specifically, there is limited research in which perceptions of the specific practices that define both autonomy supportive and thwarting practices have been examined simultaneously within an authentic classroom environment to uncover differential links with students' motivation and engagement during class. Those studies that have explored this issue have generally relied on cross-sectional designs (e.g., Assor et al., 2002; Haerens et al., 2015; but see Jang et al., 2016 for a longitudinal example of the differentiated effects of perceived teacher autonomy support and control broadly defined rather than on specific practices within each category). Thus, current research is limited to the extent that a

single retrospective survey of students' experiences is limited for drawing conclusions about the predictive role of student perceptions of teachers' various daily practices in their daily motivation and engagement in the classroom. However, teachers' autonomy relevant practices in the classroom and students' perceptions of those practices is likely to vary from one class day to the next and even minor variation is likely to change a student's daily motivation and engagement relative to his or her own typical level, though research has yet to explore this possibility. It is important to note that questions regarding the extent to which perceptions of daily teacher practices predict students' daily functioning in the classroom are distinct from questions regarding the relationships between perceptions of teachers' average practice across a semester, school year, or other period of time and students' summative motivation and engagement. Moreover, the former can only be addressed by research that monitors perceptions of teacher practice and students' classroom functioning across multiple days.

With these conceptual and methodological considerations in mind, the current investigation utilized a 6-week diary study that included regular student reports to examine differential associations between perceived autonomy supporting and thwarting teacher practices and high school science students' autonomous motivation and engagement and controlled motivation and disaffection in the classroom. Prior research suggested that we should expect perceived practices routinely identified as supportive, such as providing choices, considering students' interest and preferences in classroom activities, giving rationales about importance or usefulness, and providing opportunities for and being responsive to questions, to be strong predictors of autonomous motivation and engagement. In contrast, perceived practices routinely identified as autonomy thwarting, such as controlling messages, suppression of student perspectives, and use of uninteresting activities, were expected to be strong predictors of controlled motivation and disaffection.

Reciprocal Effects Between Perceived Teacher Practice and Students' Motivation and Engagement

According to self-determination theory, one of the primary antecedents to students' daily motivation and engagement in the classroom is expected to be their perceptions of teachers' autonomy relevant practices (e.g., Cheon & Reeve, 2013; Jang, Kim, & Reeve, 2016). However, one limitation of prior research focused on autonomy relevant teacher practices is that it has infrequently considered the extent to which student motivation and engagement may also influence perceptions of teacher practice or even objective teacher practice. Although infrequently examined, some research has suggested that teachers' respond to students' engagement. For example, Skinner and Belmont (1993) revealed in path analyses that student behavioral engagement measured in the fall was associated with the teachers' autonomy supportive behavior with students during the subsequent spring. Pelletier, Seguin-Levesque, and Legault (2002) found that when teachers perceived their students to be autonomously motivated, they were more autonomy supportive in their teaching. Jang, Kim, and Reeve (2016) found that Korean high school students' disaffection (but not engagement) predicted increases in both students' perceptions of teacher control and decreases in perceptions of teacher autonomy support over the course of a semester.

With this prior cross-sectional and longitudinal research as a base, we predicted that science students' motivation and engagement would also predict their perceptions of teachers' autonomy relevant practice on a day-to-day basis. However, in line with the dual process model, we expected to observe a differential pattern of effects across various forms of motivation and engagement. Namely, we predicted that students' perceptions that teachers' engaged in autonomy supportive practices would increase on days when students experienced autonomous motivation and engagement. Likewise, we expected that students' controlled motivation and disaffection would predict an increase in perceptions that teachers engaged in thwarting practices.

Interactions Between Perceptions of Autonomy Supportive and Thwarting Practices

Given that relatively few studies have simultaneously examined both autonomy supportive and thwarting practices, the extent to which autonomy supportive and thwarting teaching practices may yield stronger or weaker effects depending on the extent to which they are perceived to be administered in combination is a related matter that has been left unaddressed in the literature. Self-determination theory and research generally suggest that autonomy support is most effective when a cluster of supportive practices is administered together (e.g., Deci, Eghrari, Patrick, & Leone, 1994; Patall et al., 2013). However, what happens to students' motivation and engagement in a real classroom where teachers are likely to use both supportive and thwarting practices to some extent? Research indicates that perceptions of autonomy supportive and thwarting practices are distinct dimensions of teaching that are only weakly correlated (e.g., Assor et al., 2002; Haerens et al., 2015). Thus, autonomy supportive and thwarting teaching practices are likely to vary in the extent to which they are perceived to co-occur, though we know nothing about their interactive effects on students' motivation and engagement. We would argue that this issue of interaction between perceived autonomy supporting and thwarting practice is likely to be particularly relevant to the science classroom, given the emphasis in science on both discovery and innovation, as well as using established rigorous methods, rules, and procedures (e.g., Allchin, 1999). The precarious balance between these core values in science might make it particularly likely for students to perceive science teachers as using both autonomy supportive and controlling practices during the same class.

With that in mind, our predictions about how students' perceptions of supportive teaching practices might interact with thwarting practices was relatively uncertain. One possibility is that the perceived presence of thwarting teaching practices might dampen any desirable effects of perceptions of supportive practice on autonomous motivation and engagement. That is, in the context of thwarting practices, students may experience autonomy support as ineffective or insincere, limiting its functional significance for enhancing autonomous motivation and engagement. Likewise, the perceptions of autonomy support may dampen undesirable effects of thwarting practices, allaying the association between perceived thwarting practices and students' controlled motivation and disaffection.

Alternatively, a contrast interactive pattern might emerge such that in the perceived presence of teachers' thwarting practices, perceived autonomy support may predict autonomous motivation and engagement even more strongly. That is, the co-occurrence

and contrast of autonomy supportive and thwarting teacher practices may lead students to more fully appreciate the value of supportive practices and experience them as even more motivationally supportive. Likewise, thwarting practices may seem even more controlling when they are perceived to co-occur with and can be contrasted against supportive practices, bolstering the undesirable association between thwarting practices and controlled motivation and disaffection. The present investigation allowed us to test these competing hypotheses.

The Present Investigation

The aim of the present study was to test a set of theory-based hypotheses regarding the association between daily student perceptions of autonomy relevant teaching with various forms of motivation and engagement, while addressing some of the limitations in prior research by using diary methods. Given self-determination theory's assumption that it is students' subjective experiences of teachers' practice, rather than some objective reality of teacher practice, that ultimately determines students' motivation and engagement, we focused on students' perceptions of autonomy relevant teaching in the current investigation. We hypothesized that daily student perceptions of supportive practices would positively predict daily autonomous motivation and engagement in the classroom, even after controlling for the outcome on the prior class session. In contrast, we expected that daily student perceptions of autonomy thwarting teaching would yield fewer or weaker associations with those adaptive outcomes. Rather, we expected thwarting practices to be the strongest positive predictors of daily controlled motivation and disaffection in the classroom. We also expected to observe reciprocal effects from students' daily motivation and engagement to perceptions of teacher practice mimicking the same differential patterns of effect. Given the various possibilities for the patterns of interaction between students' perceptions of autonomy supportive and thwarting teaching practices and the lack of theory and prior research to guide our predictions, we made no hypotheses regarding how autonomy supportive and thwarting teaching practices may interact in their prediction of students' motivation and engagement. Finally, to strengthen confidence in the findings, we explored these hypotheses after controlling for a variety of student and classroom characteristics (e.g., students' sex, ethnicity, free or reduced price lunch eligibility, age, and prior course grade, as well as classroom content difficulty, cohort, Title I status, and teacher years of experience), because prior research has suggested that these student and classroom factors may influence students' engagement and perceptions of the environment (e.g., Clotfelter, Ladd, & Vigdor, 2010; Eccles et al., 1993; Murdock, 1999; Solomon, Battistich, & Hom, 1996), particularly within the science domain (e.g., Patall, Vasquez, Steingut, Trimble, & Pituch, 2015; Sinatra, Heddy, & Lombardi, 2015).

Overall, we expected the current study to extend evidence related to autonomy relevant teaching by contextualizing the research within an authentic high school science classroom and providing an opportunity to examine the unique, reciprocal, and interactive daily effects involving perceptions of various autonomy supportive and thwarting practices and students' daily autonomous and controlled motivation, engagement, and disaffection in the classroom. Going beyond the existing research, the current design allowed us to examine the extent to which daily variations in students' perceptions of teaching practice (or motivation and engagement) was associated with corresponding changes in students' motivation and engagement outcomes (or perceptions of teacher practice) above or below their personal baselines for engagement and motivation (or perceptions of teacher practice). We felt that this level of specificity in context, predictors, outcomes, and timing would provide the best foundation for understanding how students' experiences of teacher practice shape their motivation and engagement.

Method

Participants

There were 208 urban and suburban high school science students (13 to 18 years of age; 54% female; 68% ethnic minority; at least 43% eligible for free or reduced lunch) from 41 science classrooms across eight public high schools in the southwest region of the United States participated in this diary study. Student participants were asked to provide reports of their experiences after every science class during a 6-week instructional unit between January 2013 and May, 2014 (2,176 total reports across all students).

Every classroom was led by a different science teacher. The number of students participating in the study from each class ranged from three to six. Approximately 56% of students were enrolled in a grade-level biology, physics, or chemistry course and 44% were enrolled in an advanced biology or chemistry course or a specialty topic science course (anatomy, environmental systems, engineering, or aquatic science). Thirty-two percent of the students across these classes were White, while 42% were Hispanic/Latino, 10% were Black, 2% were Asian, and 14% were of mixed ethnicities or another ethnicity. Two students did not share their ethnicity. Forty-two percent of students were in the 9th grade, 24% were 10th graders, 17% were 11th graders, and 17% were 12th graders. The mean grade point average (GPA) at the start of the study was 2.92 (SD = 0.96; minimum = 0.82, maximum = 4.0) on a 4-point scale.

Regarding the representativeness of our sample, the urban district from which students were drawn serves a population of students in which 52% are economically disadvantaged, 67% are Hispanic or Black, and 26% are White. The suburban district from which students were drawn serves a population of students in which 22% are economically disadvantaged, 28% are Hispanic or Black, and 63% are White. Thus, a comparison of the racial and economic make-up of our student sample across both districts' student demographics suggests that we successfully recruited a student sample that was representative of the student populations being served at these eight schools.

Participation was voluntary and students under the age of 18 secured parental permission to participate. In recruiting students, the goal was to randomly select five student participants from each class among students who volunteered to participate. In the majority of classrooms (35 of 41), at least five students volunteered to participate and students were randomly selected in cases where more than five volunteers were available. In the majority of classes, approximately five to eight students volunteered to participate. Five students participated in each of 25 classes and six

students participated in each of 10 other classes. In some classes, less than five students volunteered. Four students participated in each of 5 classes and in one class just three students participated. Despite randomly selecting among volunteers in classes in which we were able, given that participation was contingent on volunteering and a limited number of students in each class volunteered, this sample should not be mistaken for a true random sample and should be considered a convenience sample. Students were paid $5 for every survey completed and received a $50 bonus for completing all reports for which they did not have an excused absence from class.

Teachers' years of experience ranged from 0 to 40 (M = 10.40, SD = 9.85). Teachers were 25 to 66 years of age (M = 38.12, SD = 12.49). The majority of teachers (30) were White and female (30). One teacher was Black, three were Asian, three were Hispanic/Latino and four were of mixed ethnicities or another ethnicity. Teachers received $50 for their participation in the study and schools received $100 for each participating teacher.

Procedure

Recruitment of participants for this study occurred in stages. Teachers were recruited in group information sessions after obtaining permission from the two school districts, as well as individual high school principals, vice principals, and science chairs at each of the eight schools. During the teacher information session, teachers were informed that the purpose of the study was to examine the relationship between students' experiences in the classroom and their motivation and engagement. The diary methods involved in the study were also explained to teachers. Participating teachers selected the course that would participate in the study and the instructional unit during which the study occurred in consultation with the research team. Teachers were encouraged to view participation in the study as an educational experience, because they would be provided information about students' motivation and engagement at the end of the study and all the information collected as part of the study was confidential. With that in mind, the research team encouraged teachers to select their most typical course for participation that suited the study best for scheduling reasons and contained a diverse group of students. The research team discouraged teachers from selecting a course because they felt it was the one in which they or their students would perform best (or worst). Across all schools, approximately 50% of recruited teachers expressed willingness to participate and approximately 40% actually participated in the study.

Student participants were recruited via in-person classroom visits in which the study was described and a parent information letter and consent documents in both English and Spanish were distributed. Students were asked to return signed consent documents in a sealed envelope to a box located at the main office of the school.

Upon recruitment and selection, participating students first met with a member of the research team to learn about their responsibilities as a participant, as well as to receive and set-up an Apple iPod touch used to complete surveys for the duration of the diary study. During this initial meeting, student participants practiced using the iPod by completing a short background survey regarding their age, grade level, sex, ethnicity, eligibility for free or reduced lunch at school based on U.S. government policy, school GPA, and course grade for the most recent instructional unit. In addition, this initial meeting was used to establish the student's school and personal schedule and determine the ideal time for the student to receive and complete daily reports.

On every class day of the 6-week instructional unit, students were emailed during their first free period (i.e., noninstructional time) after the class session with a survey asking them to respond to questions about their teachers' practices and their experiences of motivation and engagement in class. All questionnaires were programmed using Qualtrics and completed by students online using the Apple iPod touch provided by the researchers. All classes met on a block schedule, approximately every other school day. The number of report opportunities varied depending on the class and number of class sessions that occurred in the particular 6-week instructional unit. The number of scheduled class sessions ranged between 11 and 17, with classes having between 8 and 17 opportunities to report on class experiences as a result of various disruptions to class sessions (Median = 14). Daily report surveys remained available for students to complete until the next class session began. The number of reports that student participants completed across the instructional unit ranged from 1 to 17 (M = 10, SD = 3.77; Mode = 10). Only one student completed just one report and this student's responses could not be used in the analyses.

This design of repeatedly sampling students' daily perceptions of the classroom environment and experiences of motivation and engagement during class over the course of a 6-week unit allowed us to confidently examine (given the many repeated reports) ongoing within-person covariation between daily perceptions of teacher practices and experiences of motivation and engagement. That is, repeatedly sampling of participants allowed us to explore whether, for example, daily variations in perceived practices were associated with corresponding variation in motivation and engagement above or below a student's personal baseline level. Given the intense nature of drawing repeated reports from student participants over a 6-week period, we necessarily limited the number of participating students in each classroom. Restricting the number of participants from each class naturally limited the conclusions we might draw about the perceptions of the classroom from the students in the class as a whole. However, our focus was on understanding within-person (daily) variability in perceived practice, motivation, and engagement rather than variability between students or classrooms of students.

Measures

Motivation. Students' daily motivation in science class was assessed with 12 items we adapted for our use in a daily diary design from the Academic Self-Regulation Questionnaire (Ryan & Connell, 1989). This measure assessed student motivation toward education in terms of why they worked on course work, participated in science class, and tried to do well on assignments for science class *that day*. Students indicated the extent to which they engaged in each activity for intrinsic ("because it was interesting and enjoyable"), identified ("because it was important and valuable to me"), introjected ("to avoid feeling guilty or anxious"), or extrinsic reasons ("because the situation forced me to"). Students rated the extent to which they agreed with each item on a 5-point Likert scale ranging from not at all *true* (1) to *extremely true* (5). The validity and reliability of the multiscale measure for cross-

sectional research has been established in previous studies (Ryan & Connell, 1989). However, given that we adapted and shortened the measures to use them in a diary design, we conducted factor and reliability analyses to confirm that these adapted measures were appropriate for our daily diary context.

To assess the factorial validity of daily measures of motivation, we conducted a multilevel confirmatory factor analysis (ML-CFA) with four factors at both the day and student levels in Mplus 6.12. Parameters were estimated using a maximum likelihood estimation procedure (i.e., MLR) that is robust to violations of both the assumptions of normality and independence of observations, and provides for optimal parameter estimates when data are missing at random. We examined both day- and student-level (by computing the mean across class days for each student) factor structures, as factor structures are not always identical at different levels of analysis. Given the complexity of modeling a three-level exploratory factor structure and because we had just 43 classes at Level 3, we used the TYPE = COMPLEX TWO LEVEL command in Mplus to adjust *SE*s and χ^2 tests of model fit, accounting for the clustering at the classroom level (Level 3). To obtain proper estimates at each level, we followed standard multilevel modeling practices and used group-mean centering for the items at both the day and student levels using the student as the group for the lowest level and the class as the group for the student level. A well-fitting model was defined by a comparative fit index (CFI) of approximately .95, root mean square error of approximation (RMSEA) around .05, square root mean square residual (SRMR) around .08, and factor loadings >.40 (Kline, 2010). Items were allowed to load only on their target factor (i.e., intrinsic, identified, introjected, or external) and factors were allowed to correlate.

Inspection of model fit indices (CFI) > .99, RMSEA = .011, and a SRMR = .018 for the day level and .023 for the student level) indicated that the model fit the data well (Kline, 2010). Factor loadings (i.e., standardized regression coefficients) at both levels suggested that items loaded sufficiently (>.65) onto their respective factors. The correlation between intrinsic motivation and identified regulation factors was .58 and the correlation between introjected and external regulation factors was .51. The correlation between other pairs of factors ranged between −.11 and .28.

For the purposes of this study and given that our hypotheses distinguished primarily between more and less autonomous forms of motivation, we created a composite autonomous motivation variable by averaging daily intrinsic motivation and identified regulation scales (mean daily $\alpha = .92$) and a composite controlled motivation variable by averaging daily introjected and external regulation scales (mean daily $\alpha = .90$). This approach is consistent with the use of this scale in cross-sectional and experimental research (e.g., Sheldon, Ryan, Deci, & Kasser, 2004; Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004).

Engagement. Students' daily engagement in science class was assessed with 20 items we adapted for our use in a daily diary design from the Engagement versus Disaffection with Learning Student Report (Furrer & Skinner, 2003; Skinner & Belmont, 1993; Skinner et al., 2009) and the Metacognitive Strategies Questionnaire (Wolters, 2004). The Engagement versus Disaffection with Learning Student Report contains four scales from which we selected items and adapted for the daily context: behavioral engagement (3 items; e.g., "I worked as hard as I can in science class today"; "I paid attention today in science class"), emotional engagement (4 items, e.g., "I felt interested today in science class"; "I enjoyed science class today"), behavioral disaffection (3 items, e.g., "Today in science class I just did enough to get by"; "When I was in science class today, I was thinking about other things"), and emotional disaffection (6 items; e.g., "When I was in science class today, I felt bad"; "I felt unhappy in science class today"). Four items measuring learning strategies adapted from the Metacognitive Strategies Questionnaire were used to assess students' cognitive engagement in science class (e.g., "I tried to connect what I was learning in science class today with my own experiences"; "I tried to make different ideas fit together and make sense in science class today"). For all engagement items, students rated the extent to which they agreed with each item on a 5-point Likert scale ranging from *not at all true* (1) to *extremely true* (5). The validity and reliability of all engagement scales for cross-sectional research have been established in previous studies (Furrer & Skinner, 2003; Wolters, 2004). Again, given that we adapted and shortened the measures to use them in a diary design, we conducted factor and reliability analyses to confirm that these adapted measures were appropriate for our daily diary context.

We conducted a multilevel confirmatory factor analyses (ML-CFA) using MLR to examine the six factor structure at both day and student levels and the TYPE = COMPLEX TWO LEVEL syntax in Mplus to account for clustering at the classroom level. Again, items were group-mean centered for both the day and student levels using the student as the group for the lowest level and the class as the group for the student level. Items were allowed to load only on their target factor and factors were allowed to correlate. Inspection of fit indices for the model (CFI = .92, RMSEA = .03, and SRMR < .04 both for day and student levels) indicated that the model fit the data adequately (Kline, 2010). Factor loadings at both levels suggested that items loaded sufficiently (>.40) onto their respective factors.

Again, for the purposes of this study and given that our hypotheses distinguished primarily between engagement and disaffection, we created a composite engagement variable by averaging daily behavioral, emotional, and cognitive engagement scales (mean daily $\alpha = .89$) and a composite disaffection variable by averaging daily behavioral and emotional disaffection scales (mean daily $\alpha = .87$). Sizable correlations between factors supported this approach. The correlation between behavioral, emotional and cognitive engagement factors ranged between .42 and .66. The correlation between behavioral and emotional disaffection factors was .44. Moreover, using aggregated measures across types of engagement (and motivation) outcomes was an appealing approach given that it limited the number of statistical tests conducted and yielded excellent reliability characteristics.

Daily teacher practices. Students' perceptions of the extent to which their teachers used practices intended to support or thwart autonomy on a given class day was assessed with a measure designed explicitly for use in this diary study (see Appendix for final set of items) and based on prior measures used in cross-sectional research (Patall et al., 2013; as well as Assor et al., 2002, 2005; Connell, 1990; Katz, Kaplan, & Gueta, 2009; Reeve & Jang, 2006; Reeve et al., 2004; Reeve, 2006; Wellborn & Connell, 1987; Belmont, Skinner, Wellborn, & Connell, 1992). Twenty-six items assessed perceptions of five supportive daily practices and three thwarting daily practices hypothesized to be related to autonomy

need satisfaction and motivation based on prior research (e.g., Assor et al., 2002, 2005; Deci, Eghrari, Patrick, & Leone, 1994; Patall et al., 2013; Reeve, 2009; Reeve & Jang, 2006). Supportive practices included (a) provision of choices (3 items; e.g., "My teacher provided options for the kinds of assignments or activities I could do today"), (b) opportunities for students to work in their own way (3 items; e.g., "My teacher allowed me to choose how to do my work in the classroom today"), (c) consideration for student opinions, preferences, and interests (5 items; e.g., "My teacher structured class activities today around my interests"), (d) rationales regarding the usefulness and importance of course material (4 items; e.g., "My teacher explained how what we were learning today is important"), and (e) student question opportunities and responding (3 items; e.g., "My teacher acknowledged and responded to my questions in class today"). Thwarting teacher practices included (a) controlling messages (3 items; e.g., "My teacher was strict about me doing everything in his or her way today"), (b) suppression of student perspectives (3 items; e.g., "My teacher stopped me from expressing my opinions in class today"), and (c) uninteresting activities (2 items; e.g., "My teacher forced me to do uninteresting activities in class today"). Students rated the extent to which they agreed with each item on a 5-point Likert scale ranging from *not at all true* (1) to *extremely true* (5).

To assess the factorial validity of daily measures of perceived teacher practices, we conducted two multilevel exploratory factor analyses (Roesch et al., 2010) using the oblique geomin rotation and MLR in Mplus 6.12 to examine both day and student level factor structures. The first analysis included perceived supportive teacher practices and the second analysis included perceived thwarting teacher practices. These models varied in the number of factors specified at each level of the nested data structure (from 1 to 7 factors). Again, we used the TYPE = COMPLEX TWO LEVEL command in Mplus to account for the clustering at the classroom level and group-mean items for both the day and student levels using the student as the group for the lowest level and the class as the group for the student level. To determine the best-fitting model, we used a DCFI of .01 or greater as our model selection criterion (Cheung & Rensvold, 2002).

The results of ML-EFAs of these 26 items plus five additional items reflective of perceived teacher practices unrelated to this investigation supported a six factor structure for supports CFI = .98, RMSEA = .018, SRMR (day/student) = .007/.009) and a three factor structure for thwarts (CFI = .997, RMSEA = .012, SRMR (day/student) = .006/.003). All items loaded sufficiently (>.40) on the intended factor as expected with minimal cross-loadings, with the caveat that perceptions of provision of choice items and opportunities for students to work in their own way items loading on a single factor rather than two separate factors. Several items were retained only at the student level. These included one item assessing the provision of choice, two items assessing consideration for student interests and preferences, and one item related to teacher question opportunities and responding.

Supportive teacher practice factors were positively intercorrelated with small to medium correlations at the day (.14–.45) level. Likewise, thwarting teacher practice factors were positively intercorrelated with moderate correlations at the day level (.36–.46). In summary, perceived teacher practice variables were intercorrelated, but distinct (model fit deteriorated significantly if fewer or more factors were extracted). Correlations between all perceived practices are reported in the results.

Scale scores for each perceived teacher practice were calculated by taking the mean of all items loading above .40 on each factor. When factor analyses suggested that a slightly different version of a scale should be used at day versus student levels, we computed multiple versions of the scale to be used at the appropriate level. However, for the purposes of this investigation, we used only day level scales, though results were nearly identical using either version of the scales. For perceived supportive practices, the mean daily alpha was .83 for the provision of choice scale (5 items), .87 for consideration for student interests and preferences (3 items), .86 for rational provision (4 items), and .80 for question opportunities (2 items). For perceived thwarting teacher practices, the mean daily alpha was .67 for the controlling messages scale (3 items), .81 for suppression of student perspectives (3 items), and .82 for use of uninteresting activities (2 items).

Multilevel Analyses

We tested our main hypotheses about the relationships between students' daily perceptions of teacher practices and their daily experiences of motivation and engagement with a series of three-level (day, student, and class) regressions where the intercept was allowed to vary randomly using the Mixed procedure in SPSS 21. In line with recommendations from experts on conducting intensive longitudinal designs (e.g., Bolger & Laurenceau, 2013), we used hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) for our primary tests because it appropriately addressed nonindependence of observations and the hierarchically nested design of our data set in which lower level units (i.e., days) were nested within a second higher level unit (i.e., students) and students were nested within a third higher level unit (i.e., classrooms). HLM treats student and classroom as a random rather than a fixed effect, thereby permitting generalizations of the findings to a wider population.

For all multilevel models, at Level 1 (day level) we included time and the outcome reported on the previous day, in addition to daily perceived practice predictors (or daily motivation and engagement for our reciprocal models). We constructed the time variable by consecutively numbering each class session during the unit starting with zero. We opted to use class session as the time metric, as opposed to calendar days or school days elapsed, given Kim-Spoon and Grimm's (2016) recommendation to consider the dominant reasons for why changes in the outcome might occur when selecting a time metric. In our investigation, the dominant reason student motivation and engagement in science class was expected to vary is because of their experiences during science class sessions. The prior class session's value for the outcome was entered to control for possible carryover effects from one class day to the next (e.g., see Reis, Sheldon, Gable, Roscoe, & Ryan, 2000 for an example of this strategy). To minimize missing data, the most recent prior day of reporting was carried forward to the next available day of reporting for the purposes of creating lagged variables. Including the prior class session's outcome value as a predictor allowed us to predict day-to-day change in the outcome rather than sheer level (Cohen & Cohen, 1982) as a function of students' perceptions of teacher practices reported on the same class day as the outcome.

At Level 2 (student level), we included several control variables: student sex (0 = male, 1 = female), student ethnicity (0 = White or Asian, 1 = Black, Hispanic/Latino, or other ethnic minority), students' free or reduced price lunch eligibility (0 = not eligible, 1 = eligible), students' age, and students' course grade for the prior unit in all models. At Level 3 (class level), we included variables representing whether the class was advanced or grade typical (0 = grade typical, 1 = advanced), the cohort school year in which students participated in the study (0 = 2012–2013, 1 = 2013–2014), whether the classroom was in a school that had title I status or not (0 = no title I status, 1 = title I status), and teacher years of experience in all models.

To decompose within-student (day) effects from between-student effects, daily perceived practice predictors (or daily motivation and engagement predictors in reciprocal models) were student-mean centered (around each student's own average score). Time and the value of the outcome variable from the prior class session were grand-mean centered because they were simply control variables in these models, as were the nine other covariates. To treat missing data, we used a maximum likelihood estimation procedure with robust estimates of *SE*s (REML). Because adjacent residuals in repeated measures data may be correlated across measurement occasions, we specified an AR(1) correlated error structure (Bolger & Laurenceau, 2013).

Results

Preliminary Analyses

To gauge within-person variation from one class session to the next during the 6-week instructional unit compared with the variation across students and classrooms (over days), we computed variance partition coefficients (VPC; Goldstein, 2011) and intraclass correlation coefficients (ICC; Kreft & De Leeuw, 1998) for each perceived teacher practice and student self-reported engagement and motivation variable (see Table 1). VPCs suggested that between 39 and 56% of the variance in perceived teacher practices was at the day level, with a similar amount of variance at the student level and less variability at the classroom level. Similarly, VPCs suggested that between 32 and 40% of the variance in motivation and engagement was at the day level, with slightly more variance at the student level and more limited variability observed at the classroom level. Results suggested that there was a substantial proportion of daily variation in students' perceptions of their teachers' practices and their motivation and engagement over the course of the unit. Moreover, though variation at the class level was relatively small, it was still sufficiently large to warrant including a variance component at the class level (see Kreft & de Leeuw, 1998; Moerbeek, 2004).

Correlations Between Perceived Practices, Motivation, and Engagement

First, we computed correlations among the perceived daily teacher practices, engagement, and motivation variables (see Table 2). For these correlations, we group-mean centered variables using the student as the group to disentangle within-student from between-student relationships. As expected, all perceived daily practices hypothesized to be supportive of autonomy were positively correlated. Likewise, all the perceived daily practices hypothesized to be thwarting of autonomy were positively correlated. Of note, correlations among practices were modest, ranging from .12 to .33. As for correlations between supporting and thwarting practices, correlations generally hovered close to zero, ranging from −.18 to .08. Taken together, the modest values among perceived practices correlations suggest that it would be informative to investigate the effects of the seven teacher practices separately.

Table 1
Variance Partition Coefficients (VPC) and Intraclass Correlation Coefficients (ICC)

	Day level	Student level		Class level
Variable	VPC	VPC	ICC	VPC/ICC
Daily teacher practices				
Choice	.52	.40	.47	.08
Interests	.43	.43	.57	.14
Rationales	.39	.45	.61	.16
Questions	.56	.37	.44	.07
Controlling messages	.48	.48	.52	.04
Suppression	.40	.57	.60	.03
Uninteresting activities	.42	.53	.58	.05
Daily engagement and motivation				
Engagement	.34	.55	.66	.11
Disaffection	.40	.56	.60	.04
Autonomous motivation	.34	.52	.66	.15
Controlled motivation	.32	.62	.68	.05

Note. Level 1 (daily reports) n = 2,026 to 2,176 reports. Level 2 (students) n = 208. Level 3 (classes) n = 41. Calculation of the VPC and ICC is identical at the highest level of any model.

In line with our hypotheses, the four supportive daily practices were significant and positively correlated with daily engagement and autonomous motivation in class, while correlations with daily disaffection and controlled motivation hovered close to zero. Likewise, the three thwarting daily practices were significant and positively correlated with daily disaffection and controlled motivation in class. Correlations with daily engagement and autonomous motivation hovered close to zero for daily controlling messages and suppression of student perspectives, but were significant and positive for daily use of uninteresting activities.

We also computed correlations between students' perceptions of practices, motivation, and engagement aggregated across the unit and various student and classroom characteristics (see Table 3). There were a number of instances in which student and classroom characteristics (sex, ethnicity, age, free or reduced price lunch eligibility, prior course grade, type of course, teacher years of experience, school title I status, and cohort) significantly correlated with perceived teacher practices, students' motivation, or students' engagement. As such, we opted to include these variables as covariates in subsequent multilevel models.

Daily Perceived Practices as Predictors of Daily Motivation and Engagement

Next, hypotheses about the extent to which students' daily perceptions of teacher practices predict their daily experiences

Table 2
Means, SDs, and Correlations Among Daily Variables

Variable	*M* (*SD*)	1	2	3	4	5	6	7	8	9	10
1. Choice	2.47 (.96)	—									
2. Interests	2.05 (1.03)	**.33**	—								
3. Rationales	2.86 (1.05)	**.13**	**.17**	—							
4. Questions	3.63 (1.06)	**.18**	**.12**	**.22**	—						
5. Controlling messages	2.41 (.91)	*−.04*	*.04*	**.08**	**.07**	—					
6. Suppression	1.55 (.83)	*.01*	**.09**	*−.02*	**−.18**	**.25**	—				
7. Uninteresting activities	1.97 (1.08)	*−.04*	*−.07*	*−.03*	**−.13**	**.21**	**.26**	—			
8. Engagement	3.05 (.84)	**.18**	**.28**	**.31**	**.34**	**.06**	*−.04*	**−.21**	—		
9. Disaffection	1.91 (.75)	*−.01*	−.05	**−.05**	**−.09**	**.19**	**.26**	**.38**	**−.30**	—	
10. Autonomous motivation	2.94 (1.07)	**.15**	**.30**	**.26**	**.30**	*.04*	*−.03*	**−.22**	**.68**	**−.26**	—
11. Controlled motivation	2.30 (1.02)	*.02*	*.03*	*.03*	**.06**	**.25**	**.15**	**.22**	**.09**	**.27**	**.13**

Note. n = 1,998 to 2,176 reports. Correlations are computed with group-mean centered daily variables using student as the group. Italicized correlations are not significant. All other correlations (bolded) are $p < .05$.

of motivation and engagement were tested with four random intercept only three-level (day, student, and class) regressions that included all seven daily teacher practices. Results (see Table 4) largely confirmed our hypotheses that perceptions of daily autonomy supportive practices would primarily predict daily autonomous motivation and engagement, while perceptions of daily thwarting practices would primarily predict daily controlled motivation and disaffection, controlling for both time and the outcome on the prior class session, as well as a number of student and class characteristics. Specifically, all four perceived daily supportive practices (provision of choices, consideration for student interests, rationales about importance or usefulness, and question opportunities) predicted an increase in daily engagement since the prior class session, and all perceived daily supportive practices but the provision of choice predicted an increase in daily autonomous motivation from the previous class session. One perceived daily thwarting practice, daily use of uninteresting activities, also predicted a decrease in autonomous motivation and engagement since the prior class session.

Table 3
Means, SDs, and Correlations Among Student Demographic Variables and Aggregated Daily Variables

Variable	*M* (*SD*)	Sex	Ethnicity	Age	Free lunch	Prior unit course grade	Advanced course	Cohort	Title I school	Teacher experience
Sex	.54 (.50)	—								
Ethnicity	.63 (.48)	*.05*	—							
Age	15.54 (1.26)	*−.02*	*.10*	—						
Free lunch	.43 (.50)	*.02*	**.47**	*.009*	—					
Prior unit course grade	82.21 (18.10)	*−.05*	**−.16**	*.07*	**−.18**	—				
Advanced class	.44 (.50)	*−.04*	*.01*	*−.12*	*−.10*	*.06*	—			
Cohort	.58 (.49)	*.02*	*.06*	**.21**	**.16**	*.02*	**−.22**	—		
Title I school	.46 (.50)	*−.04*	**.37**	**.15**	**.43**	*−.01*	*.02*	**.22**	—	
Teacher experience	10.45 (9.53)	*.009*	**−.23**	**.18**	**−.23**	*−.03*	*−.13*	**.19**	**−.32**	—
Aggregated daily student perceived teacher practices, motivation, and engagement										
Choice	2.47 (.70)	*−.09*	*.11*	*.12*	**.23**	*−.04*	*−.01*	**.18**	**.18**	**−.15**
Interests	2.03 (.81)	*−.08*	*.11*	*.05*	**.23**	*−.05*	*−.07*	**.26**	**.26**	*−.11*
Rationales	2.85 (.85)	**−.14**	**.15**	*.05*	*.05*	*.03*	*.06*	*.14*	**.21**	*−.04*
Questions	3.60 (.75)	*−.01*	*−.13*	**.19**	**−.23**	**.28**	**.15**	*.08*	*−.05*	*.08*
Controlling messages	2.41 (.68)	*−.10*	*−.11*	*−.10*	*.07*	*−.07*	*.02*	*−.01*	*−.03*	*.01*
Suppression	1.55 (.67)	*−.07*	*.03*	**−.19**	*.12*	**−.16**	*−.07*	*.03*	*.02*	*−.05*
Uninteresting activities	1.96 (.86)	*−.05*	**−.24**	**−.18**	*−.06*	**−.13**	*−.06*	*−.06*	**−.21**	*.06*
Engagement	3.04 (.70)	**−.18**	*.09*	*.09*	*−.02*	*.13*	*.07*	*.07*	*.13*	*−.12*
Disaffection	1.91 (.61)	*.06*	**−.14**	*−.10*	*.03*	**−.16**	*−.08*	*.08*	*−.08*	*.03*
Autonomous motivation	2.92 (.89)	*−.12*	*.06*	*.08*	*−.06*	*.10*	*.06*	*.08*	*.09*	*−.08*
Controlled motivation	2.28 (.85)	*.06*	**−.19**	**−.15**	**−.19**	*−.01*	*.05*	*.05*	**−.23**	*.03*

Note. n = 199 to 208 students. Perceived teacher practice, engagement, and motivation variables were aggregated across class sessions for individual students. For student sex, 0 = male and 1 = female. For ethnicity, 0 = White or Asian and 1 = Black, Hispanic/Latino, or other ethnic minority. For free lunch, 0 = not eligible for free/reduced price lunch and 1 = eligible for free/reduced price lunch. For class type, 0 = grade typical class and 1 = advanced class. For cohort, 0 = 2012–2013 school year and 1 = 2013–2014 school year. For Title I school, 0 = not Title I status and 1 = Title I status. Students' age and prior course grade were measured continuously. Teacher experience was measured continuously as the number of years teachers' had been professionally teaching. Italicized correlations are not significant. All other correlations (bolded) are $p < .05$.

Table 4
Multilevel Regressions With Student Perceptions of Daily Teacher Practices Predicting Daily Student Motivation and Engagement

Fixed effects	Engagement		Disaffection		Autonomous motivation		Controlled motivation	
	b(*SE*)	β	*b*(*SE*)	β	*b*(*SE*)	β	*b*(*SE*)	β
Class level								
Intercept	3.06 (.05)		1.92 (.04)		2.96 (.07)		2.32 (.05)	
Advanced class	.07 (.10)	.04	−.09 (.09)	−.06	.06 (.15)	.03	.01 (.09)	.005
Cohort	−.08 (.15)	−.04	−.10 (.13)	−.05	−.12 (.22)	−.05	−.09 (.14)	−.04
Title I	.10 (.13)	.06	.01 (.11)	.004	.11 (.18)	.05	−.20 (.12)	−.10
Teacher experience	−.01 (.01)	−.07	−.002 (.005)	−.04	−.01 (.01)	−.09	−.01 (.01)	−.07
Student level								
Sex	−.17 (.08)	−.10*	.07 (.08)	.04	−.15 (.11)	−.07	.09 (.09)	.04
Ethnicity	.11 (.10)	.06	−.24 (.09)	−.16***	.08 (.13)	.04	−.10 (.11)	−.05
Age	.04 (.04)	.05	−.03 (.03)	−.05	.05 (.06)	.06	−.04 (.04)	−.05
Free/reduced lunch	−.08 (.10)	−.05	.13 (.09)	.09	−.12 (.13)	−.05	−.13 (.11)	−.06
Prior unit course grade	.002 (.002)	.04	−.005 (.002)	−.11*	.001 (.003)	.02	−.001 (.003)	−.01
Day level								
Choice	.07 (.02)	.05***	.01 (.02)	.01	.03 (.02)	.02	.02 (.02)	.01
Interests	.14 (.02)	.11***	−.01 (.02)	−.01	.23 (.02)	.13***	.01 (.02)	.01
Rationales	.16 (.02)	.12***	−.01 (.02)	−.01	.14 (.02)	.08***	.01 (.02)	.01
Questions	.13 (.02)	.11***	−.05 (.02)	−.05***	.14 (.02)	.10***	.02 (.02)	.02
Controlling messages	.003 (.02)	.002	.06 (.02)	.05***	.001 (.02)	.0004	.15 (.02)	.08***
Suppression	.002 (.02)	.001	.10 (.03)	.07***	.03 (.03)	.01	.04 (.03)	.02
Uninteresting activities	−.11 (.02)	−.09***	.21 (.02)	.19***	−.16 (.02)	−.10***	.14 (.02)	.09***
Time	−.005 (.003)	−.03*	−.0002 (.003)	.002	−.009 (.003)	−.04**	.001 (.003)	.005
Lagged outcome	.16 (.02)	.17***	.15 (.02)	.15***	.17 (.02)	.17***	.29 (.02)	.29***

Random effects	Variance	*SE*	Variance	*SE*	Variance	*SE*	Variance	*SE*
Class (Level 3) intercept	.03	.03	.02	.02	.10	.05	.001	.02
Student (Level 2) intercept	.28***	.04	.22***	.04	.45***	.07	.36***	.06
Day (Level 1)								
Residual	.16***	.006	.19***	.007	.27***	.01	.24***	.009
Autocorrelation	.01	.05	−.02	.06	−.03	.05	−.06	.06
Model Fit Statistics								
AIC	2249.14		2507.40		3155.64		2926.11	
BIC	2270.73		2529.00		3177.24		2947.71	

Note. Level 1 (daily reports) n = 1,652 to 1,654 reports. Level 2 (students) n = 190. Level 3 (classes) n = 41. The "time" variable reflects the day of reporting across the 6 week instructional unit. The "lagged outcome" variable reflects the prior class session's value for the outcome. For student sex, 0 = male and 1 = female. For student ethnicity, 0 = White or Asian and 1 = Black, Hispanic/Latino, or other ethnic minority. For free and reduced lunch status, 0 = not eligible for free/reduced lunch and 1 = eligible for free/reduced lunch. For advanced class, 0 = grade typical class and 1 = advanced class. For cohort, 0 = 2012–2013 school year and 1 = 2013–2014 school year. For Title I school, 0 = not Title I status and 1 = Title I status. b = unstandardized regression coefficient. β = standardized regression coefficient. Standardized estimates were computed using the following formula (Hox, 2010): $\beta = (b^{*}sdx)/sdy$. AIC = Akaike's Information Criterion; BIC = Schwarz's Bayesian Criterion.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

In contrast, student perceptions for most of the daily supportive practices did not predict either daily disaffection or controlled motivation. Rather, all three perceived daily thwarting practices (controlling messages, suppression of student perspectives, and use of uninteresting activities) predicted an increase in disaffection since the prior class session, and two of three, controlling messages and use of uninteresting activities, predicted an increase in daily controlled motivation. Only one perceived daily supportive practice, question opportunities, predicted a decrease in daily disaffection and none predicted a change from the previous class session in daily controlled motivation.

For the covariates, sex predicted engagement such that female students reported lower engagement across the 6 weeks than male students. Ethnicity and prior unit course grade negatively predicted disaffection. That is, Black and Hispanic students and students with higher prior grades reported experiencing less daily disaffection across the 6 weeks compared with their White or Asian and lower achieving counterparts.

Daily Motivation and Engagement as Predictors of Composite Perceived Practices

To explore the extent to which students' daily experiences of motivation and engagement predicted perceptions of teacher practices we conducted two random intercept only three-level (day, student, and class) regressions. For this analysis, we created a composite autonomy supporting practices variable to serve as the outcome in one model by taking the mean of the four perceived supportive practices (mean daily α = .89) and a composite autonomy thwarting practices variable for the outcome in the second model by taking the mean of the three perceived thwarting practices (mean daily α = .83). For each multilevel model, at Level 1 (day level) we included time, daily autonomous motivation, controlled motivation, engagement, disaffection, and the outcome reported on the previous day. At Level 2 and 3 (student and class level), we included the same set of nine control variables as in previous models. As described previously, within-student (day)

effects were student-mean centered and covariates were grand mean centered.

Results (see Table 5) were consistent with our expectations. Students' daily autonomous motivation and engagement predicted greater perceptions that teachers' engaged in autonomy supportive practices the same day over and above perceptions of teacher autonomy support during the prior class session, time, and a variety of student and classroom characteristic covariates. Likewise, students' controlled motivation and disaffection predicted greater perceptions that teachers' engaged in autonomy thwarting practices over and above perceptions of thwarts during the prior class session, time, and a variety of student and classroom characteristic covariates. The size of these effects were similar to those observed for the effects of perceived daily practices on students' daily motivation and engagement. In addition, smaller effects emerged for students' daily disaffection on perceived daily autonomy supporting practices and students' daily autonomous motivation on perceived autonomy thwarting practices. Specifically, on days when students experienced greater disaffection, they perceived slightly greater autonomy support from their teachers that same day, even after accounting for their level of perceived autonomy support in the prior class session. Moreover, on days when students experienced greater daily autonomous motivation, they perceived slightly less autonomy thwarting practices that same day, controlling for their perceptions of autonomy thwarting during the prior class session. Results suggest that students' experiences of motivation and engagement reciprocally influence perceptions of teachers' practices, such that when students are motivated for autonomous reasons and remain behaviorally, emotionally, and cognitively engaged, teachers are perceived to respond in kind with practices that further support that motivation and engagement. Encouragingly, when students reported being particularly disengaged, they also perceived teachers as providing autonomy support, which may reverse such disengagement. How-

Table 5
Multilevel Regressions With Daily Student Motivation and Engagement Predicting Composite Perceived Teacher Practices

Fixed effects	Autonomy supports		Autonomy thwarts	
	b(*SE*)	β	*b*(*SE*)	β
Class level				
Intercept	2.77 (.05)		1.98 (.04)	
Advanced class	.05 (.09)	.03	−.02 (.08)	−.01
Cohort	−.19 (.13)	−.11	−.02 (.11)	−.01
Title I	.16 (.11)	.10	−.07 (.10)	−.05
Teacher experience	−.005 (.005)	−.06	−.001 (.005)	−.02
Student level				
Sex	−.06 (.07)	−.04	−.08 (.07)	−.05
Ethnicity	−.06 (.08)	−.04	−.17 (.09)	−.11
Age	.04 (.03)	.06	−.07 (.03)	−.11
Free/reduced lunch	.12 (.09)	.08	.14 (.09)	.09
Prior unit course grade	.00 (.002)	.00	−.004 (.002)	−.08
Day level				
Autonomous motivation	.13 (.02)	.10***	−.05 (.02)	−.04*
Controlled motivation	−.03 (.02)	−.02	.17 (.02)	.12***
Engagement	.32 (.03)	.19***	.04 (.03)	.02
Disaffection	.09 (.03)	.05**	.26 (.02)	.16***
Time	.006 (.003)	.04*	.005 (.002)	.03**
Lagged outcome	.20 (.02)	.19***	.24 (.02)	.24***
Random effects	**Variance**	***SE***	**Variance**	***SE***
Class (Level 3) intercept	.04	.02	.006	.01
Student (Level 2) intercept	.18***	.03	.21***	.03
Day (Level 1)				
Residual	.23***	.009	.16***	.006
Autocorrelation	−.06	.06	−.13**	.006
Model Fit Statistics				
AIC	3026.81		2361.10	
BIC	3048.81		2383.10	

Note. Level 1 (daily reports) $n = 1{,}826$ reports. Level 2 (students) $n = 191$. Level 3 (classes) $n = 41$. The "time" variable reflects the day of reporting across the 6 week instructional unit. The "lagged outcome" variable reflects the prior class session's value for the outcome. For student sex, 0 = male and 1 = female. For student ethnicity, 0 = White or Asian and 1 = Black, Hispanic/Latino, or other ethnic minority. For free and reduced lunch status, 0 = not eligible for free/reduced lunch and 1 = eligible for free/reduced lunch. For advanced class, 0 = grade typical class and 1 = advanced class. For cohort, 0 = 2012–2013 school year and 1 = 2013–2014 school year. For Title I school, 0 = not Title I status and 1 = Title I status. b = unstandardized regression coefficient. β = standardized regression coefficient. Standardized estimates were computed using the following formula (Hox, 2010): $\beta = (b^{*}sdx)/sdy$. AIC = Akaike's Information Criterion; BIC = Schwarz's Bayesian Criterion.
$^{*} p < .05$. $^{**} p < .01$. $^{***} p < .001$.

ever, when students' motivation is controlled and they are disengaged in class, teachers are perceived to respond in kind with controlling strategies.

Interactions Between Composite Perceived Autonomy Supportive and Thwarting Practices

Finally, to address our research question regarding the interaction between perceptions of autonomy supporting and thwarting practices, we estimated four three-level random intercept only regressions that each included an interaction term between the two clusters of perceived daily practices. For this analysis, we again used the composite autonomy supporting practices variable and the composite autonomy thwarting practices variable. These models were similar to those previously described, except that these models each included only the two composite daily practice variables and their interaction, along with time, the lagged outcome covariate, and other student and classroom characteristic covariates. A model was estimated for each motivation and engagement outcome (engagement, disaffection, autonomous motivation, and controlled motivation).

There was a significant interaction between perceived daily autonomy supportive and thwarting practices for autonomous motivation, in addition to significant main effects of both (see Table 6). To get a better sense of this interaction, we conducted simple slope analyses that tested the relation between perceived daily supportive practice and autonomous motivation at 1 *SD* above and below the mean of thwarting practices. Likewise, we tested the relation between perceived daily thwarting practice and autonomous motivation at 1 *SD* above and below the mean of supportive practices. Simple slope analyses revealed that perceived daily supportive practice predicted an increase in autonomous motivation since the prior class session to a greater degree when daily thwarting practices were perceived to also be high (1 *SD* above the mean; $\beta = .20$, $p < .001$) compared with low (1 *SD* below the mean; $\beta = .15$, $p < .001$). Moreover, perceived daily thwarting practice predicted a decrease in autonomous motivation from the prior class session when daily supporting practices were perceived to be low (1 *SD* below the mean; $\beta = -.07$, $p < .001$), but not when daily supporting practices were perceived to be high (1 *SD* below the mean; $\beta = -.02$, $p = .28$). There were no interactions between perceived daily supporting and thwarting found for engagement, disaffection, or controlled motivation. Results suggest that the student perceptions of teachers' supporting their autonomy has a particularly strong relationship with their daily autonomous motivation when contrasted against thwarting practices perceived on the same day. Likewise, any undesirable effect of students' perceptions that their teachers are using autonomy thwarting practices on their daily autonomous motivation was mitigated when students also perceived their teachers to be engaging in supportive practices on the same day.

Table 6
Multilevel Regressions With Composite Perceived Teacher Practices and Their Interaction Predicting Autonomous Motivation

Fixed effects	*b*(*SE*)	β
Class level		
Intercept	2.94 (.07)	
Advanced class	.08 (.14)	.04
Cohort	−.09 (.20)	−.04
Title I	.10 (.17)	.05
Teacher experience	−.01 (.01)	−.06
Student level		
Sex	−.14 (.10)	−.07
Ethnicity	.09 (.12)	.04
Age	.03 (.05)	.04
Free/reduced lunch	−.11 (.13)	−.05
Prior unit course grade	.001 (.003)	.02
Day level		
Daily supports	.36 (.03)	.17***
Daily thwarts	−.11 (.03)	−.05***
Supports × Thwarts	.13 (.05)	.04**
Time	−.01 (.003)	−.04**
Lagged outcome	.21 (.02)	.21***

Random effects	Variance	*SE*
Class (Level 3) intercept	.10	.05
Student (Level 2) intercept	.39***	.07
Day level (Level 1)		
Residual	.33***	.01
Autocorrelation	−.04	.06
Model Fit Statistics		
AIC	3729.49	
BIC	3751.48	

Note. Level 1 (daily reports) $n = 1{,}820$ reports. Level 2 (students) $n = 191$. Level 3 (classes) $n = 41$. The "time" variable reflects the day of reporting across the 6 week instructional unit. The "lagged outcome" variable reflects the prior class session's value for the outcome. For student sex, 0 = male and 1 = female. For student ethnicity, 0 = White or Asian and 1 = Black, Hispanic/Latino, or other ethnic minority. For free and reduced lunch eligibility, 0 = not eligible free/reduced lunch and 1 = eligible free/reduced lunch. For advanced class, 0 = grade typical class and 1 = advanced class. For cohort, 0 = 2012–2013 school year and 1 = 2013–2014 school year. For Title I school, 0 = not Title I status and 1 = Title I status. *b* = unstandardized regression coefficient. β = standardized regression coefficient. Standardized estimates were computed using the following formula (Hox, 2010): $\beta = (b^{*}sdx)/sdy$.
** $p < .01$. *** $p < .001$.

Discussion

The present investigation examined the role of various perceived autonomy relevant teaching strategies in students' daily autonomous motivation, controlled motivation, engagement, and disaffection in authentic high school science classes, as well as reciprocal relationships among these variables. We used a diary method to track students' daily perceptions of teacher practices and experiences during science class over a 6-week instructional unit. We also explored how perceived strategies routinely identified as autonomy supporting or thwarting interact and whether the presence of one type of perceived practice moderates the relation of the other with students' motivation and engagement during class.

Fit of Data to Theoretical Predictions

Overall, the patterns of results supported our hypotheses and were consistent with the dual process model within self-determination theory (Jang et al., 2016). We found the expected differentiated effects in which changes in autonomous motivation and engagement were predicted primarily by daily perceptions of

autonomy supportive teacher practices, while changes in controlled motivation and disaffection were predicted primarily by daily perceptions of thwarting practices. More specifically, daily student perceptions that teachers' considered their preferences and interests, provided rationales about the importance of usefulness of course activities, and provided opportunities for and responses to questions consistently predicted increases in both autonomous motivation and engagement since the prior class session, and perceived choice opportunities predicted increases in engagement. In contrast, perceptions of these daily supportive practices generally did not predict controlled motivation and disaffection, with the one exception being that daily perceptions that teachers provided question opportunities negatively predicted disaffection. Rather, student perceptions of thwarting practices consistently predicted controlled motivation and disaffection. Specifically, controlling messages and use of uninteresting activities predicted an increase in both controlled motivation and disaffection since the last class session and suppression of student perceptions predicted an increase in disaffection. Among these thwarting practices, only daily perceptions that teachers' used uninteresting activities appeared to be pervasively detrimental, predicting a decrease in both daily engagement and autonomous motivation, in addition to predicting an increase in controlled motivation and disaffection. However, both controlling message sand suppression of students' perspectives were unrelated to engagement or autonomous motivation.

Regarding reciprocal effects, students' motivation and engagement also predicted changes in their perceptions of their teachers' autonomy relevant practice largely in the expected differentiated pattern. Namely, an increase in perceived autonomy support was predicted primarily by students' daily autonomous motivation and engagement, and to a lesser extent by daily disaffection. In contrast, an increase in perceived autonomy thwarting was predicted primarily by controlled motivation and disaffection, and negatively predicted by autonomous motivation to a lesser extent. One surprising finding regarding reciprocal effects was that students' disaffection predicted an increase in perceptions that teachers' engaged in autonomy supportive practices. This particular finding is somewhat inconsistent with prior traditional (nondaily diary) longitudinal evidence suggesting that disengagement predicts less autonomy support (e.g., Jang et al., 2016). Although surprising in the context of previous findings, we find this quite encouraging as it suggests that on days when students are actively disengaged during class, they perceive their teachers to react by increasing their support for autonomy during that same class session (presumably in an attempt to elicit engagement from students). We also note that we observed the expected relationship between students' autonomous motivation and engagement to perceived autonomy support on a *daily basis* even though some prior traditional longitudinal research examining reciprocal effects predicted by the dual process model (e.g., Jang et al., 2016) did not observe this relationship. Finally, it is also worth noting that the magnitude of effects in both directions were quite similar,[1] leading us to conclude that the students' experience of motivation and engagement may play an equally important role in the perceptions of the classroom environment as the classroom environment plays in students' experiences of motivation and engagement.

Taken together, evidence provided in this investigation is largely consistent with prior cross-sectional and longitudinal evidence and extends it by demonstrating the utility of the dual process model for day-to-day reciprocal links between students' perceptions of their teachers' autonomy-relevant practice, motivation, and engagement (e.g., Assor et al., 2002; Haerens et al., 2015; Jang et al., 2016). That is, the pattern of results suggests that there are largely divergent pathways to various aspects of students' functioning in the classroom. Students are likely to experience heightened behavioral, emotional, and cognitive engagement, as well as internal forms of motivation that spring from interest, enjoyment, and value on days when they perceive their teachers to use autonomy supportive practices like rationales, activities that consider students' interests, and questions, and to some extent, choices. However, the absence of these daily practices does not generally lead to students' disaffection and controlled motivation in class. Rather, it is when students perceive teachers to use explicitly controlling practices—controlling messages, suppression of student perspectives, and activities that seem uninteresting or meaningless—that students become behaviorally and emotionally disengaged and pursue school tasks for more external reasons. Likewise, students' behavioral, affective, and cognitive experiences predict their perceptions of the classroom environment (and possibly teachers' actual behavior). On days when students experience autonomous motivation and engagement, their perceptions that teachers are supportive of their autonomy increase. In contrast, on days when students experience controlled motivation and engagement in class, they perceive their teachers to be more controlling.

Interactions Between Perceived Supports and Thwarts

With the basic pattern of relationships between perceived teacher practices and students' motivation and engagement established, it was also clear that the interaction between perceived autonomy supportive and thwarting practices was somewhat complex. Given that science emphasizes both discovery and using established, rigorous procedures, there is likely to be many opportunities for both supporting autonomy (e.g., "design your own experiment on something related to what we have been studying today that interests you") and controlling behavior (e.g., "this is how you *need* to conduct this experiment if you want it to work") in science courses. With that in mind, our results suggested that we may not need to be quite so worried about students' perceiving their teachers to engage in autonomy thwarting practices on a given day as long as they also perceive teachers to engage in autonomy supportive practices. We found that perceived supportive practices predicted a greater increase in autonomous motivation on days when thwarting practices were perceived to be high compared with low. Likewise, students' perceptions that their teachers used thwarting practices only predicted a decrease in autonomous motivation on days when they perceived supportive

[1] To explicitly compare the magnitude of effects across reciprocal effects, we conducted additional multilevel models including the aggregated perceived autonomy supporting practice and perceived autonomy thwarting practice as predictors of each form of motivation and engagement. Autonomy support predicted autonomous motivation and engagement most strongly (β = .17 and .19, $ps < .001$) and controlled motivation and disaffection to a lesser extent or not at all (β = .03 and −.01, ps = .03 and .23). Autonomy thwarting predicted controlled motivation and disaffection most strongly (β = .15 and .23, $ps < .001$) and autonomous motivation and engagement to a lesser extent (β = −.04 and −,04, $ps < .005$ and .001).

practices to be low, but not on days when they were perceived supports to be high.

We found this interaction only for autonomous motivation. As such, results about the differential effects of these predictors are limited in scope. Nonetheless, these results suggest that the perceived contrast between autonomy supports and thwarts may have the desirable effect of heightening the significance of autonomy support for enhancing students' autonomous motivation. That is, in comparison with a controlling strategy that a student might have recently perceived in class, autonomy supportive strategies are perceived to be particularly supportive of a student's interests, values, and drive to engage in some class activity for internal reasons. The combination of perceived autonomy supporting and thwarting may also have the converse desirable effect of students being less sensitive to perceived controlling practices as long as they are accompanied by supportive practices. Perhaps in the context of perceiving teachers to use practices that support autonomy, teachers controlling practices are experienced as providing structure and organization, rather than attempts to control behavior and thwart students' autonomy. Despite these findings, we would not encourage teachers to intentionally use controlling practices, particularly given our findings that perceived daily thwarts clearly predicted daily disaffection and we found no evidence that perceived supports could mitigate that association. Likewise, there was no evidence that perceived supports and thwarts interacted to influence engagement and only limited evidence of interaction for controlled motivation, which we discuss next in reference to perceived suppression of student perspectives.

The Conundrum of Choice and Suppression

Another surprising finding was that daily perceptions of choice opportunities predicted increases in daily engagement, but not autonomous motivation and similarly, daily perceptions that teachers' suppressed student perspectives predicted an increase in daily disaffection, but not controlled motivation. To better understand these null findings, we conducted a number of exploratory analyses (a) examining the effects of perceived practices after decomposing the daily motivation outcomes into their constituents and (b) examining interactions involving these two particular practices and each of the other practices.

First, these exploratory multilevel model analyses revealed that students' daily perceptions that teachers provided choices predicted intrinsic motivation ($\beta = .04$, $p = .02$), but had no relationship with identified motivation ($\beta = -.0004$, $p = .98$). This finding suggests that choice provision is an autonomy supportive practice that is particularly predictive of forms of motivation based in positive emotions (i.e., interest and enjoyment) rather than value. This is consistent with prior research suggesting that choice provision is most strongly related to intrinsic motivation and less strongly related to motivation focused on the importance or value of the activity (e.g., Patall, Cooper, & Wynn, 2010; Patall et al., 2013).

Second, an exploratory multilevel model analyses also revealed that perceived choice provision interacted with perceptions of a number of other practices that changed its daily relationship with autonomous motivation. Specifically, daily perceptions that teachers' provided choices interacted with three other practices, perceived daily question opportunities ($\beta = .06$, $p < .001$), controlling messages ($\beta = .04$, $p < .009$), and use of uninteresting activities ($\beta = .03$, $p < .03$). Simple slope analyses suggested that perceptions of greater daily choice provision predicted greater autonomous motivation when opportunities to ask questions, controlling messages, or use of uninteresting activities were also perceived to be high (1 *SD* above the mean; βs = .07, .05, and .04, $ps < .001$, .01, and .02), but not when they were perceived to be low (1 *SD* below the mean; βs = −.04, −.02, and −.02; $ps = .06$, .30, and .41). Results suggest that students' perception that their teachers provided choices on a given day is specifically related to autonomous motivation during class when bolstered by the presence of another supportive practice (daily question opportunities) or contrasted against a thwarting practice (daily controlling message and uninteresting activities) on the same day.

One way to interpret this finding is to first note that, at times, choices can be overwhelming rather than motivating for students (e.g., Iyengar & Lepper, 2000; Patall, Cooper, & Robinson, 2008; Schwartz & Ward, 2004). However, when accompanied by another support that also serves to provide some structure (question opportunities), the motivating function of choosing can be revealed. That is, when students are provided with choices but are not allowed to ask questions about those choices or the activity, the choice might seem more arbitrary and less important, or students may lack confidence to make the "right" choice without the necessary information. If, on the other hand, students are provided with the opportunity to ask questions about their choice and the task, choosing may be more likely to be experienced as strategic, personal, and effective. Controlling messages may also be experienced similarly as a form of structure that can support the motivational benefits of choosing when the two are provided in combination. It is worth noting that this interpretation is consistent with research suggesting that students' motivation thrives after choosing in contexts in which they feel competent, but deteriorates after choosing if they do not feel competent (i.e., Patall, Sylvester, & Han, 2014).

Theoretically, choice is presumed to enhance the experiences of autonomy by allowing individuals to express the self and act in accordance with their personal preferences and interests (e.g., Katz & Assor, 2007; Patall et al., 2008; Ryan & Deci, 2000). Accordingly, researchers have long noted the possibility that providing choices may be particularly useful in the context of boring rather than interesting tasks because there is more opportunity to improve the task by incorporating personal preferences and interests in the context of a motivationally deprived task (e.g., Patall et al., 2013, 2010; Sansone, Weir, Harpster, & Morgan, 1992; Tafarodi, Milne, & Smith, 1999). In contrast, when a task is already interesting and autonomy-supportive by its very nature, choosing becomes an unnecessary expenditure of decision-making effort that may even diminish autonomous motivation. In fact, recent laboratory-based experiments have demonstrated that college students reported enhanced interest, perceived competence, value, and liking for a reading comprehension task after choosing aspects of the task only when the task was boring, but not when it was interesting (e.g., Patall et al., 2013). This investigation is in line with those findings, suggesting that within the science classroom, perceiving the opportunity to make choices about learning tasks and classroom activities may enhance autonomous motivation most in the context of activities that are perceived to be particularly uninteresting.

A final exploratory multilevel analysis revealed that perceived suppression interacted with perceptions of other practices that changed its daily relationship with controlled motivation. Specifically, daily perceptions that teachers' suppressed student perspectives interacted with two other practices, perceived choice provision ($\beta = .10, p < .001$) and question opportunities ($\beta = -.04, p < .02$). Perceptions that teachers' suppressed student perspectives during the class session predicted students' greater controlled motivation when opportunities to ask questions during class were perceived to be low (1 *SD* below the mean; $\beta = .05, p < .02$), but not when they were perceived to be high (1 *SD* above the mean; $\beta = -.02, p = .37$). Perceptions that teachers' suppressed student perspectives during class also predicted greater controlled motivation when the provision of choice was perceived to be high during the class session (1 *SD* above the mean, $\beta = .09, p < .001$). However, when daily perceptions of choice provision were low (1 *SD* below the mean), perceived suppression during the class negatively predicted students' controlled motivation ($\beta = -.06, p < .003$). Results suggest that the relationship between daily suppression and students' controlled motivation depends on the perception of other practices during the same class session, with the perception of question opportunities mitigating the undesirable effect of perceived suppression increasing controlled motivation, and the perception of choice opportunities magnifying that effect. The latter finding again highlights the very mixed benefits and detriments of having choices. Choices can often be experienced as overwhelming by students. When combined with the perception that teachers will not allow students to express their opinions, preferences, and feelings, the experience of being controlled and behaving merely to obtain rewards or avoid undesirable consequences is likely to be particularly robust.

Limitations and Implications for Future Research

Given the potential practical implications of understanding the links between teachers' practices and students' motivation, engagement, and achievement, it would seem imperative that future research replicate and extend the findings of the current investigation. Strengths of the current investigation include the simultaneous focus on various perceived autonomy supportive and thwarting practices, the intensive longitudinal design that allowed us to examine the extent to which daily variations in students' perceptions of teacher practice was associated with corresponding fluctuations in daily motivation and engagement in the classroom, and the fact that the study was situated within a heterogeneous set of science classrooms with students of various social, economic, and cultural backgrounds. Despite the strengths of the current design, the correlational nature of the design cannot be taken to imply causation. Consequently, findings of this investigation should be corroborated with experimental designs in authentic classroom contexts that isolate the effects of various autonomy relevant practices and allow for the interactions among them to be explored to best understand the effects of teachers' autonomy relevant practice. Thus far, intervention research focused on autonomy relevant teacher practice has generally focused on autonomy support as a whole or only one specific practice isolated from others (e.g., choice provision).

The reliance on student self-reports in the current investigation presents another significant limitation that needs to be addressed in future research. Although the focus on student perceptions of teachers' practice is reasonable given self-determination theory's assumption that it is students' subjective experiences that are the most powerful predictor of their motivation and engagement, relying exclusively on students' self-reports leaves open the possibility that response-bias and shared-method variance may influence the results. Accordingly, using independent observations of the classroom to explore the extent to which autonomy relevant teacher practice relates to students' motivation and engagement outcomes is an important next step in this scholarship, though we acknowledge that observations present their own unique set of limitations and biases. While there are examples of researchers using observation to determine teachers' autonomy supporting or thwarting practice (e.g., De Meyer et al., 2014; Reeve et al., 2004), we know of no research in which individual components of autonomy relevant practice were observed as separate coding categories and used as separate variables to predict outcomes. Given the complex dynamics that seem to play out between various autonomy relevant practices, we believe that a nuanced understanding of what makes for the best autonomy relevant teaching practice requires detailed coding at the individual teacher strategy level. This is likely to be particularly true for practices such as choice provision and suppression of student perspective, which this investigation highlighted as having particularly heterogeneous associations with other teaching practices and student outcomes.

In future research, we also encourage researchers to examine formally the extent to which need satisfaction and frustration mediates the daily relationships uncovered in this investigation. Though we selected the current set of perceived practices after reviewing previous research regarding practices that have been associated with students' perceived autonomy (e.g., Patall et al., 2013; Reeve & Jang, 2006), it is possible that various psychological processes mediate the relationships between perceived daily teaching practices and students' daily motivation and engagement. Moreover, we would be remiss if we did not point out that our list of autonomy supportive and thwarting practices is not comprehensive. Although we attempted to select the most central and promising strategies, motivation researchers have suggested a variety of additional practices, such as acknowledgment of negative affect, encouragement, perspective-taking, use of deadlines, and controlling rewards (e.g., Reeve, 2009; Reeve & Jang, 2006), that could be considered in future research focused on autonomy relevant teaching.

We also want to highlight that the nature of the design in the current investigation in which students were asked to provide reports multiple days a week for several weeks necessitated relying on a small sample of volunteers from each class. Likewise, teachers selected the participating class and were themselves volunteers. Though we attempted to recruit a diverse sample of teachers and adolescents (e.g., we randomly selected student participants among volunteers and approximately 40% of teachers across participating schools volunteered to participate), the voluntary and selective nature of the sample undoubtedly provides the opportunity for biased results that are idiosyncratic to the current sample. Future research should attempt to address this limitation with classes and samples that are randomly selected to the greatest extent possible.

Finally, although it was not the focus of this investigation, results also suggested that female students were less engaged in science class compared with male students. Given the continued

concern about engaging women in STEM (e.g., Bidwell, 2015), this finding highlights the need for future research to explore the benefits and detriments of autonomy relevant teaching practices in science domains particularly for female students and the contexts that might be most supportive of their motivation and engagement.

Conclusion

In conclusion, this investigation adds to the growing body of research exploring perceptions of autonomy relevant teaching and its reciprocal relations with adolescent students' motivation and engagement. This study goes beyond those previously conducted by using an intensive daily diary study to examine perceptions of various daily supportive *and* thwarting practices in an authentic academic classroom setting. Taken together, results suggested that students' perceptions of teachers' daily supportive and thwarting practices have distinct reciprocal relations with various aspects of students' motivation and engagement during class. While perceived supportive practices primarily predicted changes in daily autonomous motivation and engagement in class and vice versa, perceived thwarting practices primarily predicted students' daily controlled motivation and disaffection during class and vice versa. Moreover, the current investigation is the first to highlight that perceived supportive and thwarting practices interact and that the presence of both may yield benefits for students' motivation, though it is important to note that we found this interaction only for autonomous motivation. We hope that this investigation serves as a useful guide for future classroom-based theory and research focused on motivationally relevant instruction.

References

Allchin, D. (1999). Values in science: An educational perspective. *Science & Education, 8,* 1–12. http://dx.doi.org/10.1023/A:1008600230536

Archambault, I., Janosz, M., Fallu, J.-S., & Pagani, L. S. (2009). Student engagement and its relationship with early high school dropout. *Journal of Adolescence, 32,* 651–670. http://dx.doi.org/10.1016/j.adolescence.2008.06.007

Assor, A., Kaplan, H., Kanat-Maymon, Y., & Roth, G. (2005). Directly controlling teacher behaviors as predictors of poor motivation and engagement in girls and boys: The role of anger and anxiety. *Learning and Instruction, 15,* 397–413. http://dx.doi.org/10.1016/j.learninstruc.2005.07.008

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology, 72,* 261–278. http://dx.doi.org/10.1348/000709902158883

Bartholomew, K. J., Ntoumanis, N., Ryan, R. M., Bosch, J. A., & Thøgersen-Ntoumani, C. (2011). Self-determination theory and diminished functioning: The role of interpersonal control and psychological need thwarting. *Personality and Social Psychology Bulletin, 37,* 1459–1473. http://dx.doi.org/10.1177/0146167211413125

Bartos, S. A., & Lederman, N. G. (2014). Teachers' knowledge structures for nature of science and scientific inquiry: Conceptions and classroom practice. *Journal of Research in Science Teaching, 51,* 1150–1184. http://dx.doi.org/10.1002/tea.21168

Belmont, M. J., Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1992). *Teacher as social context (TASC): Two measures of teacher provision of involvement, structure, and autonomy support.* Unpublished manuscript, Graduate School of Education and Human Development, Department of Psychology, University of Rochester.

Bidwell, A. (2015, February 24). STEM workforce no more diverse than 14 years ago. *US News and World Report.* Retrieved from http://www.usnews.com

Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY: Guilford Press.

Bureau of Labor Statistics. (2011). *Occupational outlook handbook.* Washington, DC: Author.

Cheon, S. H., & Reeve, J. (2015). A classroom-based intervention to help teachers decrease students' amotivation. *Contemporary Educational Psychology, 40,* 99–111.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9,* 233–255. http://dx.doi.org/10.1207/s15328007sem0902_5

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2010). Teacher credentials and student achievement in high school: A cross-subject analysis with student fixed effects. *The Journal of Human Resources, 45,* 655–681. http://dx.doi.org/10.3368/jhr.45.3.655

Cohen, J., & Cohen, P. (1982). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Connell, J. P. (1990). Context, self and action: A motivational analysis of self-system processes across the life-span. In D. Cichhetti (Ed.), *The self in transition: From infancy to childhood* (pp. 61–97). Chicago, IL: University of Chicago Press.

Corpus, J. H., McClintic-Gilbert, M. S., & Hayenga, A. O. (2009). Within-year changes in children's intrinsic and extrinsic motivational orientations: Contextual predictors and academic outcomes. *Contemporary Educational Psychology, 34,* 154–166. http://dx.doi.org/10.1016/j.cedpsych.2009.01.001

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62,* 119–142. http://dx.doi.org/10.1111/j.1467-6494.1994.tb00797.x

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11,* 227–268. http://dx.doi.org/10.1207/S15327965PLI1104_01

De Meyer, J., Tallir, I. B., Soenens, B., Vansteenkiste, M., Aelterman, N., Van den Berghe, L., . . . Haerens, L. (2014). Does observed controlling teaching behavior relate to students' motivation in physical education? *Journal of Educational Psychology, 106,* 541–554. http://dx.doi.org/10.1037/a0034399

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. M. (1993). Development during adolescence. The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48,* 90–101. http://dx.doi.org/10.1037/0003-066X.48.2.90

Erikson, E. H. (1968). *Identity: Youth and crisis.* Oxford, England: Norton & Co.

Furrer, C., & Skinner, E. A. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology, 95,* 148–162. http://dx.doi.org/10.1037/0022-0663.95.1.148

Goldstein, H. (2011). *Multilevel Statistical Models* (4th ed.). Hoboken, NJ: Wiley.

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93,* 3–13. http://dx.doi.org/10.1037/0022-0663.93.1.3

Gottfried, A. E., Marcoulides, G. A., Gottfried, A. W., & Oliver, P. H. (2009). A latent curve model of parental motivational practices and developmental decline in math and science academic intrinsic motivation. *Journal of Educational Psychology, 101,* 729–739. http://dx.doi.org/10.1037/a0015084

Haerens, L., Aelterman, N., Vansteenkiste, M., Soenens, B., & Van Petegem, S. (2015). Do perceived autonomy-supportive and controlling teaching relate to physical education students' motivational experiences through unique pathways? Distinguishing between the bright and dark side of motivation. *Psychology of Sport and Exercise, 16,* 26–36. http://dx.doi.org/10.1016/j.psychsport.2014.08.013

Harter, S. (1981). A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components. *Developmental Psychology, 17,* 300–312. http://dx.doi.org/10.1037/0012-1649.17.3.300

Hox, J. J. (2010). *Multilevel analysis. Techniques and applications I* (2nd ed.). New York, NY: Routledge.

Hughes, J. N., Luo, W., Kwok, O.-M., & Loyd, L. K. (2008). Teacher-student support, effortful engagement, and achievement: A 3-year longitudinal study. *Journal of Educational Psychology, 100,* 1–14. http://dx.doi.org/10.1037/0022-0663.100.1.1

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79,* 995–1006. http://dx.doi.org/10.1037/0022-3514.79.6.995

Jang, H., Kim, E. J., & Reeve, J. (2012). Longitudinal test of self-determination theory's motivation mediation model in a naturally occurring classroom context. *Journal of Educational Psychology, 104,* 1175–1188. http://dx.doi.org/10.1037/a0028089

Jang, H., Kim, E. J., & Reeve, J. (2016). Why students become more engaged or more disengaged during the semester: A self-determination theory dual-process model. *Learning and Instruction, 43,* 27–38. http://dx.doi.org/10.1016/j.learninstruc.2016.01.002

Katz, I., & Assor, A. (2007). When choice motivates and when it does not. *Educational Psychology Review, 19,* 429–442. http://dx.doi.org/10.1007/s10648-006-9027-y

Katz, I., Kaplan, A., & Gueta, G. (2009). Students' needs, teachers' support, and motivation for doing homework: A cross-sectional study. *Journal of Experimental Education, 78,* 246–267. http://dx.doi.org/10.1080/00220970903292868

Kim-Spoon, J., & Grimm, K. J. (2016). Latent growth curve modeling and developmental psychology. In D. Cicchetti (Ed.), *Developmental psychopathology, Vol. 1, theory, and method* (3rd ed., pp. 986–1041). Hoboken, NJ: Wiley.

Kline, R. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press Publications.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. London, England: Sage. http://dx.doi.org/10.4135/9781849209366

Kuhn, T. S. (1962). *The structure of scientific revolutions* (1st ed.). Chicago, IL: University of Chicago Press.

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97,* 184–196. http://dx.doi.org/10.1037/0022-0663.97.2.184

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education, 95,* 877–907. http://dx.doi.org/10.1002/sce.20441

Miceli, M., & Castelfranchi, C. (2000). Nature and mechanisms of loss of motivation. *Review of General Psychology, 4,* 238–263. http://dx.doi.org/10.1037/1089-2680.4.3.238

Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39,* 129–149. http://dx.doi.org/10.1207/s15327906mbr3901_5

Murayama, K., Pekrun, R., Lichtenfeld, S., & Vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development, 84,* 1475–1490. http://dx.doi.org/10.1111/cdev.12036

Murdock, T. B. (1999). The social context of risk: Status and motivational predictors of alienation in middle school. *Journal of Educational Psychology, 91,* 62–75. http://dx.doi.org/10.1037/0022-0663.91.1.62

Otis, N., Grouzet, F. M. E., & Pelletier, L. G. (2005). Latent Motivational Change in an Academic Setting: A 3-Year Longitudinal Study. *Journal of Educational Psychology, 97,* 170–183. http://dx.doi.org/10.1037/0022-0663.97.2.170

Organization for Economic Co-operation and Development (OECD). (2006). *Evolution of student interest in science and technology studies: Policy report.* Paris, France: Organisation for Economic Co-operation and Development Global Science Forum.

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: A meta-analysis of research findings. *Psychological Bulletin, 134,* 270–300. http://dx.doi.org/10.1037/0033-2909.134.2.270

Patall, E. A., Cooper, H., & Wynn, S. R. (2010). The effectiveness and relative importance of choice in the classroom. *Journal of Educational Psychology, 102,* 896–915. http://dx.doi.org/10.1037/a0019545

Patall, E. A., Dent, A. L., Oyer, M., & Wynn, S. R. (2013). Student autonomy and course value: The unique and cumulative roles of various teacher practices. *Motivation and Emotion, 37,* 14–32. http://dx.doi.org/10.1007/s11031-012-9305-6

Patall, E. A., Sylvester, B. J., & Han, C.-W. (2014). The role of competence in the effects of choice on motivation. *Journal of Experimental Social Psychology, 50,* 27–44. http://dx.doi.org/10.1016/j.jesp.2013.09.002

Patall, E. A., Vasquez, A. C., Steingut, R. R., Trimble, S. R., & Pituch, K. A. (2016). Daily interest, engagement, and autonomy support in the high school science classroom. *Contemporary Educational Psychology, 46,* 180–194. http://dx.doi.org/10.1016/j.cedpsych.2016.06.002

Pelletier, L. G., Séguin-Lévesque, C., & Legault, L. (2002). Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors. *Journal of Educational Psychology, 94,* 186–196. http://dx.doi.org/10.1037//0022-0663.94.1.186

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Reeve, J. (2006). Teachers as facilitators: What autonomy-supportive teachers do and why their students benefit. *The Elementary School Journal, 106,* 225–236. http://dx.doi.org/10.1086/501484

Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist, 44,* 159–175. http://dx.doi.org/10.1080/00461520903028990

Reeve, J., & Jang, H. (2006). What teachers say and do to support students' autonomy during a learning activity. *Journal of Educational Psychology, 98,* 209–218. http://dx.doi.org/10.1037/0022-0663.98.1.209

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28,* 147–169. http://dx.doi.org/10.1023/B:MOEM.0000032312.95499.6f

Reis, H. T., Sheldon, K. M., Gable, S. L., Roscoe, J., & Ryan, R. M. (2000). Daily well-being: The role of autonomy, competence, and relatedness. *Personality and Social Psychology Bulletin, 26,* 419–435. http://dx.doi.org/10.1177/0146167200266002

Roesch, S. C., Aldridge, A. A., Stocking, S. N., Villodas, F., Leung, Q., Bartley, C. E., & Black, L. J. (2010). Multilevel factor analysis and structural equation modeling of daily diary coping data: Modeling trait and state variation. *Multivariate Behavioral Research, 45,* 767–789. http://dx.doi.org/10.1080/00273171.2010.519276

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57,* 749–761. http://dx.doi.org/10.1037/0022-3514.57.5.749

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55,* 68–78. http://dx.doi.org/10.1037/0003-066X.55.1.68

Sansone, C., Weir, C., Harpster, L., & Morgan, C. (1992). Once a boring task always a boring task? Interest as a self-regulatory mechanism. *Journal of Personality and Social Psychology, 63,* 379–390. http://dx.doi.org/10.1037/0022-3514.63.3.379

Schwartz, B., & Ward, A. (2004). Doing better but feeling worse: The paradox of choice. In P. A. Linley, S. Joseph, P. A. Linley, & S. Joseph (Eds.), *Positive psychology in practice* (pp. 86–104). Hoboken, NJ: Wiley http://dx.doi.org/10.1002/9780470939338.ch6

Sheldon, K. M., Ryan, R. M., Deci, E. L., & Kasser, T. (2004). The independent effects of goal contents and motives on well-being: It's both what you pursue and why you pursue it. *Personality and Social Psychology Bulletin, 30,* 475–486. http://dx.doi.org/10.1177/0146167203261883

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist, 50,* 1–13. http://dx.doi.org/10.1080/00461520.2014.1002924

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85,* 571–581. http://dx.doi.org/10.1037/0022-0663.85.4.571

Skinner, E. A., Furrer, C., Marchand, G., & Kindermann, T. A. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100,* 765–781. http://dx.doi.org/10.1037/a0012840

Skinner, E. A., Kindermann, T. A., Connell, J. P., & Wellborn, J. G. (2009). Engagement and disaffection as organizational constructs in the dynamics of motivational development. In K. R. Wenzel, A. Wigfield, K. R. Wenzel, & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 223–245). New York, NY: Routledge/Taylor & Francis Group.

Soenens, B., Sierens, E., Vansteenkiste, M., Dochy, F., & Goossens, L. (2012). Psychologically controlling teaching: Examining outcomes, antecedents, and mediators. *Journal of Educational Psychology, 104,* 108–120. http://dx.doi.org/10.1037/a0025742

Solomon, D., Battistich, V., & Hom, A. (1996). Teacher beliefs and practices in schools serving communities that differ in socioeconomic level. *Journal of Experimental Education, 64,* 327–347. http://dx.doi.org/10.1080/00220973.1996.10806602

Su, Y.-L., & Reeve, J. (2011). A meta-analysis of the effectiveness of intervention programs designed to support autonomy. *Educational Psychology Review, 23,* 159–188. http://dx.doi.org/10.1007/s10648-010-9142-7

Tafarodi, R. W., Milne, A. B., & Smith, A. J. (1999). The confidence of choice: Evidence for an augmentation effect on self-perceived performance. *Personality and Social Psychology Bulletin, 25,* 1405–1416. http://dx.doi.org/10.1177/0146167299259006

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87,* 246–260. http://dx.doi.org/10.1037/0022-3514.87.2.246

Vansteenkiste, M., Zhou, M., Lens, W., & Soenens, B. (2005). Experiences of autonomy and control among Chinese learners: Vitalizing or immobilizing? *Journal of Educational Psychology, 97,* 468–483. http://dx.doi.org/10.1037/0022-0663.97.3.468

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences, 16,* 1–12. http://dx.doi.org/10.1016/j.lindif.2005.06.004

Wellborn, J. G., & Connell, J. P. (1987). *Manual for the Rochester assessment package for schools*. Rochester, NY: University of Rochester.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39,* 1–37. http://dx.doi.org/10.1111/j.1745-3984.2002.tb01133.x

Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96,* 236–250. http://dx.doi.org/10.1037/0022-0663.96.2.236

(*Appendix follows*)

Appendix

Items from the Daily Perceived Teacher Practice Measure

Provision of Choice

My teacher allowed me to choose which questions or parts of an assignment to work on today.

My teacher provided options for the kinds of assignments or activities I could do today.

My teacher allowed me to choose how to do my work in the classroom today.

My teacher allowed me to choose how to use my time for studying and classwork today.

My teacher encouraged me to work in my own way today.

Consideration for Student Interests and Preferences

My teacher structured class activities today around my interests.

My teacher took my preferences into consideration for assignments today.

My teacher worked my interests into his or her lesson(s) today.

Rationales Identifying Usefulness, Importance, and Relevance of Activities

My teacher explained how what we were learning today is important.

My teacher demonstrated how what we were learning today is useful.

My teacher explained how the course assignments today were important.

My teacher talked about the connection between what we are studying in school today and real life.

Student Question Opportunities

My teacher provided opportunities for me to ask questions today.

My teacher acknowledged and responded to my questions in class today.

Controlling Messages

My teacher was strict about me doing everything in his or her way today.

The language my teacher used today included how I "should" or "ought" to do things.

My teacher told me to work on the assignments today because she or he said so.

Suppression of Student Perspectives and Controlling Activities

My teacher stopped me from expressing my opinions in class today.

My teacher stopped me from asking questions in class today.

My teacher prevented me from expressing complaints or talking about my negative feelings during class today.

Meaningless or Uninteresting Activities

My teacher forced me to study boring topics today.

My teacher forced me to do uninteresting activities in class today.

Received November 17, 2015
Revision received April 26, 2017
Accepted May 8, 2017 ■

Journal of Educational Psychology
2018, Vol. 110, No. 2, 289–307

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000210

The Moderating Role of Popular Peers' Achievement Goals in 5th- and 6th-Graders' Achievement-Related Friendships: A Social Network Analysis

Lydia Laninga-Wijnen
Utrecht University

Allison M. Ryan
University of Michigan

Zeena Harakeh
Utrecht University

Huiyoung Shin
Oklahoma State University

Wilma A. M. Vollebergh
Utrecht University

This research investigated whether classroom-based peer norms for achievement goals moderate friendship selection, maintenance and influence processes related to academic achievement in 46 Grade 5 and Grade 6 classrooms (N = 901, 58.7% Grade 5 students, 48.5% boys). A distinction was made between peer norms for mastery (i.e., developing competence) and performance (i.e., demonstrating competence) goals. Peer norms were measured in terms of popularity norms (the within-classroom correlation between student achievement goals and popularity) and descriptive norms (the class-level aggregated average achievement goals). As hypothesized, longitudinal social network analyses revealed that achievement goal popularity norms played a role in friendship processes, rather than achievement goal descriptive norms. Specifically, adolescents formed friendships with similarly achieving peers in classrooms with high performance goal popularity norms but not in classrooms with low performance goal popularity norms. Conversely, adolescents remained friends with similarly achieving peers in classrooms with low performance goal popularity norms but not in classrooms with high performance goal popularity norms. Furthermore, friendship influence on achievement took place in classrooms with high mastery goal popularity norms, but not in classrooms with low mastery goal popularity norms. This study indicates that friendship processes regarding achievement depend upon the extent to which certain achievement goals are made salient by virtue of their association with popularity in classrooms.

Educational Impact and Implications Statement

The salience of achievement goals is known to affect social interactions with peers around academic tasks. In contrast to prior work which has focused on how teachers make different goals salient, the current study focused on the role that popular peers play in making achievement goals salient in the classroom. Results of the current study were consistent with the idea that the goals pursued by popular peers created classroom norms that influenced friendship processes around achievement. In classrooms where popular students endorse performance goals (i.e., demonstrating competence relative to others), adolescents initially select their friends based on similarity in (high) achievement; however, these friendships among high-achieving peers do not hold over time. In classrooms where popular students endorse mastery goals (i.e., developing competence), achievement increases when adolescents have high-achieving friends but decreases when they have low-achieving friends. Teachers need to appreciate and attend to popular peers and their impact on classroom climate, friendship processes and academic achievement in early adolescence.

Keywords: achievement goals, academic achievement, friendship processes, peer norms, popularity
Supplemental materials: http://dx.doi.org/10.1037/edu0000210.supp

Academic achievement in adolescence is a crucial predictor of future educational and occupational success (Crosnoe & Benner, 2015). For better or worse, peers may provide an important developmental context for adolescent academic achievement (Rodkin & Ryan, 2012). Academic achievement may shape peer relationships through processes in which adolescents select or maintain similarly achieving others as friends; relationships, in turn, may shape individual academic achievement, through friendship socialization (i.e., influence) processes. These processes result in similarity in academic achievement among friends. However, friendship selection, maintenance and influence processes do not operate in isolation, but take place in broader peer contexts, such as classrooms and schools (Veenstra & Dijkstra, 2011), which may play a role in the direction and magnitude of these friendship dynamics. One

This article was published Online First June 22, 2017.

Lydia Laninga-Wijnen, Department of Interdisciplinary Social Sciences, Utrecht University; Allison M. Ryan, School of Education, University of Michigan; Zeena Harakeh, Department of Interdisciplinary Social Sciences, Utrecht University; Huiyoung Shin, Department of Educational Psychology, Oklahoma State University; Wilma A. M. Vollebergh, Department of Interdisciplinary Social Sciences, Utrecht University.

We thank Christian Steglich (University of Groningen) for his useful advice on our analyses.

Correspondence concerning this article should be addressed to Lydia Laninga-Wijnen, Utrecht University, Room H220, P.O. Box 80.140, 3508 TC, Utrecht. E-mail: a.m.wijnen@uu.nl

way of characterizing the broader social context in the classroom is by using the concept of *peer norms* (Dijkstra & Gest, 2015). As peer norms reflect the expected and accepted behaviors and attitudes of a social group (Shaw, 1981), they may play a role in determining whether academic achievement is a salient attribute for friendship selection, maintenance and influence processes. Therefore, the current study examined the role of peer norms in friendship processes (i.e., selection, maintenance and influence) related to adolescents' academic achievement (see Figure 1 for a conceptual model).

In the current article, we focus on *peer-perceived achievement* (or *academic reputation*; Gest, Rulison, Davidson, & Welsh, 2008) as index of academic achievement, as this has both practical and theoretical value for the current study. First, peer-perceived achievement has been shown to be a valid indicator of adolescent academic competence that is highly correlated with grade point average (GPA; correlations varying from .60 to .70; Gest et al., 2008), but that also captures unique information on how well adolescents are doing at school. Peers can be seen as expert observers and have a unique perspective on classmates' academic functioning, because their proximity to and direct interaction with classmates permit unique observations about the speed and ease (or difficulty) with which classmates finish assignments, expend effort on tasks, and give or receive help. These insights may not always be captured by tests, grade point averages or teacher ratings (Gest et al., 2008). Second, having a positive academic reputation (i.e., high peer-perceived achievement) may be associated with having academic successes recognized and remembered by peers, being approached more often for academic help (which is fruitful for one's own academic development as well), and affiliating with other classmates perceived as high-achieving (Greenwood, 1991), which in turn may have implications for friendship selection, maintenance and influence processes.

Friendship Selection, Maintenance, and Influence Processes Related to Achievement (Figures 1a and 1b)

Theoretically, selection and maintenance of friends on the basis of similarity in achievement can be explained with the similarity-attraction hypothesis (Byrne & Nelson, 1965), which states that adolescents prefer interacting with partners who maintain similar attitudes and values, as this enhances perceived trust and predictability in social interactions (Byrne & Lamberth, 1971). Friends may not only be similar in achievement due to selection or maintenance processes, but also due to socialization (i.e., influence) processes. Friends are assumed to socialize adolescents' achievement through information exchange, modeling, reinforcement of peer norms and values (Kindermann & Gest, 2009; Ryan, 2000), and peer tutoring experiences (Gest et al., 2008).

Innovative methodological advances in social network analysis allow researchers to disentangle the dynamic, reciprocal interplay of friendship selection, maintenance, and influence processes in a methodologically sound way, yielding reliable and accurate indications of the strength and direction of these processes (using stochastic actor-based models; Steglich, Snijders, & Pearson, 2010). A few previous studies have applied these statistical techniques to investigate the extent to which friendship selection and influence processes related to achievement take place, but their findings on the presence and direction of friendship processes vary considerably across and within studies; and only one study addressed friendship maintenance processes.

With regard to the presence of friendship processes, one study on high-school students (Grades 9 and 10) found that influence, maintenance and (especially) selection processes contributed to similarity in achievement among friends (Rambaran et al., 2016), whereas another study on elementary students (Grade 6) found influence but not selection processes contributed to similarity in achievement among friends (Shin & Ryan, 2014a). Furthermore, in one other previous study, the extent to which selection and influence were present varied across contexts within the study. This study of Flashman (2012) on high-school-students' academic achievement in eight schools (Grades 7 through 12) indicated that both selection and influence explained similarity in grade point averages (i.e., GPA rank) between high school friends at the two largest schools analyzed, but not at the six small, private and rural schools analyzed.

With regard to the direction of friendship processes, one study indicated that friendship selection and maintenance mainly occurred among similarly low-achieving peers (Rambaran et al., 2016), whereas the direction of friendship selection varied between

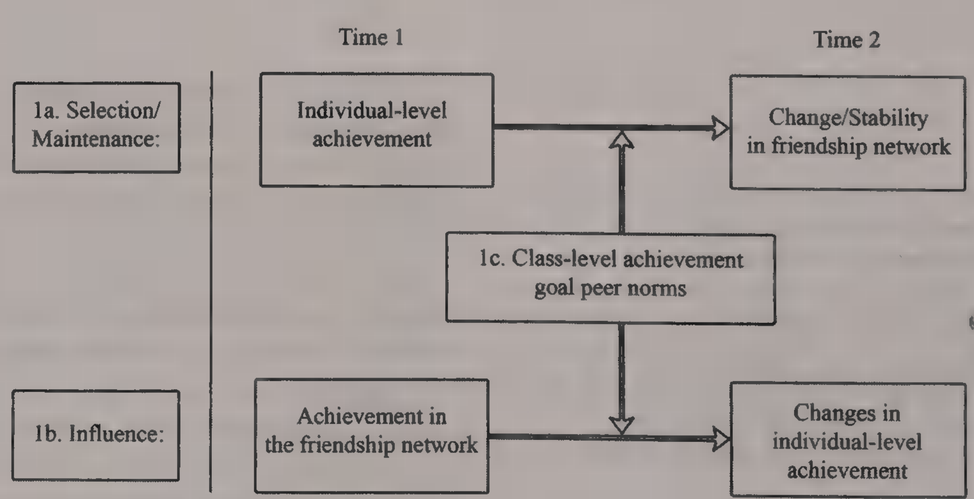

Figure 1. Conceptual model on the role of achievement goal peer norms in friendship selection, maintenance, and influence on achievement.

schools in the study of Flashman (2012). That is, in one large, public school, high-achieving students mostly formed relations with other high-achieving students, whereas in the other large, public school, similarity-based selection took place equally among low-achieving students and high-achieving students. Only one study examined the direction of friendship influence on achievement, indicating that friends influenced one another to increase rather than decrease in achievement over time (Rambaran et al., 2016). On the basis of these studies, it can be concluded that maintenance processes have been underinvestigated, and, more importantly, that the magnitude and the direction of friendship selection and influence processes varied across studies and even across different settings within the same study (i.e., larger schools compared with smaller schools; Flashman, 2012). So far, studies have only reported this variation between settings; an explanation of why selection and influence processes vary across different settings is lacking.

In the current study, we propose that one reason why friendship processes related to achievement may vary across settings is that different settings represent different peer contexts, which in turn have diverging implications for friendship processes (Kiuru et al., 2012). One way of measuring the peer context is by using the concept of peer norms which has received attention in several recent studies due to its linkages with adolescent behavior and peer relations (Dijkstra & Gest, 2015; Laninga-Wijnen et al., 2016; McCormick & Cappella, 2014; Rambaran, Dijkstra, & Stark, 2013). Peer norms represent the expected and appropriate behaviors and attitudes in a particular setting and, therefore, may determine the valence of certain behaviors for friendship selection, maintenance and influence processes (McCormick & Cappella, 2014). That is, according to social misfit theory (Wright, Giammarino, & Parad, 1986), adolescents have a tendency to conform to the peer norm in order to fit in with the expectations of the peer group and to gain acceptance and avoid rejection by their peers. When adolescents are liked in a particular setting, peers may perceive them as attractive friendship partners and, hence, these adolescents have a greater chance of being selected and maintained as friends. Furthermore, based on social identity theory (Tajfel & Turner, 1986) it could be reasoned that adolescents may be especially susceptible for friendship influence related to behaviors that are in line with the peer norm, as this yields a shared identity that provides emotional and social support, behavioral confirmation and a sense of self. Therefore, peer norms in the classroom may foster friendship selection, maintenance and influence processes, for instance related to achievement (Veenstra & Dijkstra, 2011).

Indeed, two previous studies indicated that peer norms played an important role in determining the direction and magnitude of friendship influence and selection processes related to peer-perceived aggression (Laninga-Wijnen et al., 2016) and risk attitudes (Rambaran et al., 2013). In the current study we will extend this work by examining whether classroom-based peer norms for achievement goals also play a role in friendship processes related to achievement. As detailed next, we consider achievement goal peer norms given extensive theory and research about the importance of achievement goals for academic beliefs and behaviors as well as for interpersonal relations in the classroom (Linnenbrink-Garcia & Patall, 2016; Poortvliet & Darnon, 2010; Wigfield et al., 2016).

Achievement Goal Popularity Norms and Friendship Processes (Figure 1c)

In achievement settings, two contrasting goals are often evident: mastery and performance goals (Ames, 1992; Dweck, 1986; Elliot, 2005). When mastery goals are salient in the classroom, there is a focus on developing academic competence or task mastery, whereas when performance goals are salient, there is a focus on demonstrating academic competence relative to other students, through superior performance or looking smart (Pintrich, 2000). An extensive body of research has shown that the salience of these achievement goals (due to manipulation in experiments or natural variation in classrooms) affects academic motivation and behavior (Anderman & Wolters, 2006; Linnenbrink-Garcia & Patall, 2016; Wigfield et al., 2016). Relevant to the present study, achievement goals have been found to influence social interactions with peers on academic tasks (Darnon, Dompnier, & Poortvliet, 2012; Levy-Tossman, Kaplan, & Assor, 2007; Levy-Tossman, Kaplan, & Patrick, 2004; Poortvliet & Darnon, 2010).

In the achievement goal literature, theory and research have tended to focus on how teachers make achievement goals salient in the classroom (Ames, 1992; Patrick, Mantzicopoulos, & Sears, 2012). However, teachers and peers both contribute to the classroom context (Pianta & Hamre, 2009). In the current study, we focus on how peers can make particular achievement goals salient within the classroom, as during early adolescence, students may become more likely to model behaviors after their peers and might be less likely to model parent or teacher behaviors (Cairns, Cairns, Xie, Leung, & Hearne, 1998; Galván, Spatzier, & Juvonen, 2011; Sumter, Bokhorst, Steinberg, & Westenberg, 2009).

Ushered in with the pubertal and social changes of early adolescence, youth show increased susceptibility to peer influence during this stage (Steinberg, 2007). Peers can set a norm for adolescent's academic behaviors and attitudes in the classroom (McCormick & Cappella, 2014). Yet, it is unlikely that all peers are equally influential, and during early adolescence especially popular peers may set a norm within the classroom (Rambaran et al., 2013) as there is a peak in the desire for popularity among peers during this age period (LaFontana & Cillessen, 2010). As a result, adolescents may be highly attuned to the behaviors and attitudes of popular peers, as these behaviors and attitudes are reputationally salient (reputational salience hypothesis; Hartup, 1996). This implies that these behaviors and attitudes are positively valued within a setting and an important tool for improving an adolescent's own reputation (i.e., popularity).

Popular students can make the achievement goals they endorse salient (i.e., set a norm) within the classroom via task-related messages that refer to mastery or performance goals, or via academic behaviors and endeavors (Urdan & Schoenfelder, 2006). More specifically, students are likely to voice various reasons for and reactions to their work that may refer to mastery goals or performance goals, respectively. For instance, when working on some math problems, some students might especially try to hurry and be the first to finish (performance goal), whereas others might focus on really learning the material, solving problems themselves, and not compare themselves to others (mastery goals). All of these goals could go together with visible behaviors and explicit comments (e.g., "Yeah I am first compared with all of you!" or "Yeah, I solved this problem myself!"; see Shin & Ryan, 2014b). Indeed,

numerous studies and experiments have indicated that achievement goals are outwardly exhibited and can be recognized by specific behaviors and messages referring to these goals (see for instance Darnon et al., 2012, and Poortvliet & Darnon, 2010, for an overview). In this way, students may notice the goals of popular peers.

One approach to capture the norms of popular adolescents (i.e., the popularity norm; Laninga-Wijnen et al., 2016) is by examining the within-classroom correlation between popularity and behaviors or attitudes (also referred to as *norm salience*; Henry et al., 2000; Rambaran et al., 2013). These achievement goal popularity norms, in turn, may have important implications for the coevolution of interpersonal relations and achievement within the classroom, which we describe in the following text.

Friendship Selection and Maintenance Related to Achievement

Performance goal peer norms and mastery goal peer norms can be linked with friendship selection and maintenance processes related to achievement based on social comparison theory (Festinger, 1954). In classrooms where performance goals are salient, interpersonal standards are used to define relative competence. As a result, adolescents tend to use social comparison to ensure that they did better (or not worse) than others in terms of their achievement (see Brophy, 2005, for a review). These social comparison processes may play a role in friendship selection and maintenance processes related to achievement in two ways. On the one hand, it could be hypothesized that when performance goals are salient in the classroom, students have a higher tendency to select and maintain similarly achieving peers as friends than when performance goals are less salient in the classroom. Differences in achievement to a friend can be threatening because of the emphasis on social comparison and achievement as validating one's sense of self-worth (Elliot, Murayama, & Pekrun, 2011; Festinger, 1954). When the levels of achievement are similar, comparisons would be less threatening for self-worth. Therefore, it could be expected that when popularity norms make performance goals salient, similarity-based selection and maintenance related to achievement would take place, both among low-achieving students and among high-achieving students. On the other hand, it could be hypothesized that when performance goals are salient, students have a lower tendency to select and maintain similarly achieving peers as friends because they have self-enhancement motives. More specifically, to fulfill the need of maintaining a positive self-view (i.e., self-enhancement), adolescents may have a tendency to select and maintain lower achieving peers as friends and use them as a proximal comparison standard (Régner, Escribe, & Dupeyrat, 2007) to boost their self-view with a favorable comparison.

In regards to mastery goal norms and friendship selection, two alternate hypotheses can be formulated as well. When mastery goals are salient, the focus is on personal improvement and task mastery, and not on interpersonal differences in achievement (Poortvliet & Darnon, 2010). In such a situation, achievement differences (i.e., social comparison; Festinger, 1954) among students may be less important or valuable for friendship selection and maintenance processes. Therefore, it can be hypothesized that mastery goal peer norms may not be powerful enough to break down the general tendency of selecting and maintaining similar friends (similarity-attraction hypothesis, Byrne & Lamberth, 1971), which would result in similarity-based selection and maintenance of friends irrespective of whether popularity norms make mastery goals salient. On the other hand, it can be considered that classrooms with salient mastery goals *are* characterized by social comparison, as social comparison with others can also serve the goal of self-improvement (Collins, 1996, 2000). Social comparison can be a useful learning resource for gaining accurate information for self-evaluation and acquiring information about how to improve, which are compatible with the requirements of mastery goals (Butler, 1995; Collins, 1996, 2000; Lockwood & Kunda, 1997; Régner et al., 2007). In this way, achievement may be a valuable characteristic and important indicator of competence, and students may use social comparison (i.e., upward comparison) to seek out friends they can learn from (i.e., the high-achieving students). Therefore, it could also be hypothesized that when popularity norms make mastery goals salient, friendship selection and maintenance takes place based on dissimilarity in achievement, with students selecting and maintaining higher achieving peers as friends.

Friendship Influence on Achievement

Achievement goal peer norms can be linked with friendship influence processes by social interdependence theory (Johnson & Johnson, 1989, 2005). Social interdependence exists when individual goal attainment is affected by others' actions (Johnson & Johnson, 1989, 2005). There are two types of interdependence: *Positive interdependence* refers to a situation in which there is a positive relation between goal attainments of individuals, whereas *negative interdependence* exists when individuals perceive that they can obtain their goals (only) if the other individuals with whom they are competitively linked fail to reach their goals (Deutsch, 1949, 1962; Johnson & Johnson, 1989, 2005). The extent to which a classroom is characterized by positive or negative interdependence has implications for social interactions around academic tasks (Deutsch, 1949, 1962; Roseth et al., 2008), and hence, for the magnitude and direction of friendship processes related to achievement.

In classrooms where performance goals are salient, individuals may experience negative interdependence with their classmates (also referred to as a *competitive goal structure*; Deutsch, 1949, 1962; Elliot et al., 2016), because they reach their goals when others do not reach their goals, as they aim at outperforming others (Poortvliet & Darnon, 2010). This negative interdependence may result in oppositional interaction patterns within the classroom, with individuals discouraging and obstructing each other's efforts to achieve their goals. In such a situation, individuals focus both on being productive and on preventing any other person from being more productive than themselves (Deutsch, 1949). In other words, individuals may develop an exploitation orientation toward information exchange, which reflects the incentive to profit from task-related efforts of exchange partners, paired with a reluctance to offer good or valuable information in return (Poortvliet, Janssen, Van Yperen, & Van de Vliert, 2007; Poortvliet & Darnon, 2010). Indeed, previous studies indicated that when performance goals are made salient within a setting, individuals have a reduced willingness to coordinate efforts with potential exchange partners, a reluctance to be dependent on the actions of others (for instance

with regard to asking for help; Ryan, Gheen, & Midgley, 1998; Ryan & Shim, 2012), and a reduced readiness to be influenced by exchange (see Poortvliet & Darnon, 2010). There may even be suspiciousness about exchanging information as performance goals have been linked to tactically deceiving peers in order to outperform them (Poortvliet, Anseel, Janssen, Van Yperen, & Van de Vliert, 2012). Hence, on the basis of the social interdependence theory (Deutsch, 1949, 1962) it could be hypothesized that when popularity norms make performance goals salient, productive social interactions around academic tasks are less likely, which minimizes the opportunities for friends to influence each other and become similar over time.

When mastery goals are salient in a setting, students are likely to perceive positive interdependence with fellow students (Elliot et al., 2016; Poortvliet & Darnon, 2010), as they see others as helpers in achieving their goals (Karabenick, 2003; Roussel, Elliot, & Feltman, 2011; Ryan & Shim, 2012). Positive interdependence (also referred to as a *cooperative goal structure*; Deutsch, 1949, 1962) is associated with promotive interaction, implying that individuals encourage and facilitate each other's efforts to complete tasks in order to reach the group's goals (Deutsch, 1949). Social exchanges can serve as an important means by which individuals can obtain their goal of self-improvement, which may enhance an adolescent's willingness to invest in relationship building with potential exchange partners. Indeed, previous research indicated that when mastery goals are salient, students have a higher tendency to reciprocally share valuable information, actively engage in adaptive help-seeking, have constructive discussions and collaborate on academic issues (Darnon et al., 2012; Karabenick, 2003; Ryan & Shim, 2012). Also, mastery goals have been linked to the provision of resources and effort to help team members who are apparently failing to perform well (Porter, 2005). We therefore hypothesized that in classrooms where popularity norms make mastery goals salient, the conditions and processes through which friends have the potential to influence each other are enhanced (Kindermann & Gest, 2009; Ryan, 2000), which results in more similarity among friends in achievement. More specifically, we expect that the promotive interaction patterns will result in positive friendship influence; that is, we expect that friends will influence adolescents to increase rather than to decrease in achievement over time.

Achievement Goal Descriptive Norms and Friendship Processes (Figure 1c)

Another approach to examine classroom peer norms and achievement goals is to use descriptive norms rather than popularity norms. Descriptive norms refer to the average behaviors or attitudes of all peers in a given setting, for instance a classroom (Wright et al., 1986). However, previous studies indicated that descriptive norms were not predictive of variations in friendship processes regarding peer-perceived aggression (Laninga-Wijnen et al., 2016) and risk attitudes (Rambaran et al., 2013). According to social impact theory, the strength of social forces (in this case: peer norms) is a function of the status of peers, closeness of peers, and number of peers present (Latané, 1981). Descriptive norms only represent the last, quite subtle aspect of this function and hence may not be strong enough to determine social impact (Laninga-Wijnen et al., 2016). Therefore, we do not expect that descriptive norms play a role in friendship processes related to achievement. However, given the examination of popularity and descriptive norms in relation to friendship processes is quite new, we examine both to add to the empirical evidence on this issue.

Present Study

We examined the role of achievement goal peer norms in friendship processes related to achievement (see Figure 1). We hypothesized that achievement goal popularity norms rather than achievement goal descriptive norms would play a role in friendship processes related to achievement, because popularity norms represent the behaviors and attitudes that are positively valued in classrooms (i.e., reputationally salient; Hartup, 1996), especially during early adolescence. We conducted our investigation in the context of math and science classrooms, where academic achievement is likely to be especially salient to peers. In contrast to language arts or social studies classrooms, which often emphasize writing and evaluating information that can be interpreted in different ways, math and science coursework more often involves formulas and clear-cut "right" or "wrong" answers (Franke, Kazemi, & Battey, 2007; Martin, Way, Bobis, & Anderson, 2015; Fredricks et al., 2016). Thus, it may be easier for students to garner information about their peers' performance in math and science classrooms because they can more readily compare results on assignments and tests (Stodolsky & Grossman, 1995; Wang, Fredricks, Hofkens, & Schall, 2016).

Method

Procedure and Participants

Data were collected as part of the Classroom and Peer Ecologies Project, a longitudinal study examining early adolescent social and academic adjustment in school. Schools were recruited from three school districts located in small urban communities with comparable demographics in the Midwest region of the United States. The school districts serve a sizable proportion of low-income (50% to 71%) as well as middle-income families. In these school districts the elementary schools contained students in kindergarten through Grade 5 and the middle schools contained Grades 6 through 8. All of the middle schools in these districts (N = 6) agreed to participate in the project. Two feeder elementary schools for each middle school also agreed to participate (N = 12). In the elementary schools, children were in a self-contained classroom with one teacher for the majority of the day. In the middle schools, students rotated among different teachers for their main academic subjects. However, middle school students and teachers were organized into smaller teams within their grade level, so students saw many of the same peers in their different classrooms at middle school.

To provide a common reference point across the different school settings, we focused on the classroom context in the domains of math and science (for a similar approach, see Eccles et al., 1993; Midgley, 2002). We focused on both math and science to garner a higher number of unique teachers and distinct classrooms at the middle school level than would have been possible had we exclusively focused on just math or science teachers. All math and science teachers in Grade 6 at the middle schools agreed to

participate and we chose one of their classrooms to administer surveys. For the teachers from the feeder elementary schools, we aimed to focus on math or science in equal proportions (e.g., if there were two math and two science teachers at the middle school we would focus on math class for one teacher and science class for the other teacher within each of the two feeder elementary schools). Two factors contributed to our sample having more math than science classrooms: (1) there were more Grade 6 math than science teachers in the middle schools and (2) for some elementary school teachers, science instruction was not occurring during the time frame of our study (e.g., science and social studies instruction would alternate every few weeks) and in those cases we conducted our investigation in math class.

Letters describing the project were given to all students to take home to their parents early in the school year. Eighty-four percent of the students returned permission slips granting them parental approval to participate. About 2 to 3 months into the school year, surveys were administered to students in their classrooms by two trained research assistants. Instructions and items were read aloud while students read along and responded. Survey administration was repeated about 6 months later in the spring of the school year. All classrooms did not complete all measures at Wave 2 due to timing and scheduling constraints (predominantly coming from one elementary and one middle school). The missing data included measures used in the present study and thus students from those classrooms were not included in this investigation. The total sample (N = 901 at Wave 1 and N = 859 at Wave 2) was about half female (51.5%) and ethnically diverse (36.8% African American, 46.9% European American, 7.5% Hispanic, and 8.8% other ethnic groups). Students came from 46 classrooms, each with different teachers and students (19 classrooms at the Grade 6 level, consisting of 11 math classrooms and 8 science classrooms and 27 classrooms at the Grade 5 level, consisting of 20 math classrooms and 7 science classrooms) situated within 16 schools (5 middle schools and 11 elementary schools).

Measures

Friendship networks. Adolescents' friends within classrooms were measured by asking students to nominate their friends in the classroom, further described to students as "the friends you hang around with and talk to the most." Embedded in each child's survey was a class list, and students were told they could nominate as many or as few friends as they wanted by putting a check next to names of their friends. Friendship networks were calculated for each classroom. A value of 1 equaled a given friendship nomination, whereas a value of 0 depicted an absent nomination.

Peer-perceived academic achievement. Students were asked to nominate which peers within the classroom "gets good grades." Similar to the friendship networks, students put a check next to names on a class list that followed the question. The number of nominations received were standardized by class for all participants into z scores. Because RSIENA analyses (Ripley, Snijders, Boda, Vörös, & Preciado, 2016) require ordinal categorical dependent behavior variables, these peer perceived achievement z scores were recoded into four roughly equally populated categories based on quartiles (for Wave 1: Category 1 = $z \leq -.737$; Category 2 = $-.737 < z \leq -.338$; Category 3 = $-.338 < z \leq .581$ and Category 4 = $z > .581$; for Wave 2: Category 1 = $z \leq -.748$; Category 2 = $-.748 < z \leq -.392$; Category 3 = $-.392 < z \leq .580$; and Category 4 = $z > .580$).

Achievement goal peer norms. Achievement goal popularity norms were measured at Time 1 (T1) as the within-classroom correlation between peer-nominated popularity and achievement goals (Dijkstra & Gest, 2015; Dijkstra, Lindenberg, & Veenstra, 2008; Laninga-Wijnen et al., 2016).

Peer-nominated popularity was assessed by taking the average of two items: (1) "Which students in this class do you admire most?" and (2) "Which students in your class are really cool?"; in line with Sandstrom (2011). The correlations between these two items were r = .60 and r = .70 for Waves 1 and 2, respectively (both $p < .001$). To assess the achievement goals of students, we used the Patterns of Adaptive Learning Survey (Midgley, Arunkumar, & Urdan, 1996). Mastery goals were measured with six items focusing on developing academic competence (e.g., "An important reason I do my math/ science work is because I want to improve my skills" and "An important reason I do my math/ science work is because I like to learn new things"). Performance goals were measured using five items focusing on demonstrating high academic competence relative to other students in the class (e.g., "Doing better than other students in my math/science class is important to me" and "An important reason I do my math/science work is because I want to do better than other students in my class"). Participants were asked to rate on a 5-point Likert scale, ranging from 1 (*not at all true*) to 5 (*very true*). The scales measuring achievement goals were found to be reliable in the present sample at both time points (Cronbach's α = .84 and .87 for mastery goals, and .84 and .87 for performance goals, for Waves 1 and 2, respectively). The mastery items and the performance items were averaged, to create scales for mastery goals and performance goals, respectively.

We made a distinction between three types of classrooms based on quartiles of the within-classroom correlation between popularity and achievement goals. Classrooms with low popularity norms were characterized by a correlation in the lowest quartile for performance or mastery goal popularity norms (low mastery: $r < -.13$, $N_{\text{classrooms}}$ = 11; low performance: $r < -.26$, $N_{\text{classrooms}}$ = 11). Classrooms with average popularity norms scored in the middle quartiles (25% to 75%) of achievement goal popularity norms (moderate mastery: $-.13 \leq r \leq .29$, $N_{\text{classrooms}}$ = 24; moderate performance: $-.26 \leq r \leq .11$, $N_{\text{classrooms}}$ = 24). Classrooms with high popularity norms scored in the highest quartile of achievement goal popularity norms (high mastery, $r > .29$, $N_{\text{classrooms}}$ = 11; high performance, $r > .11$; $N_{\text{classrooms}}$ = 11).

Descriptive norms were measured at T1 as the aggregated average score for mastery and performance goals, respectively, across all students in the class (Dijkstra & Gest, 2015; Rambaran et al., 2013; Laninga-Wijnen et al., 2016). We made a distinction between three types of classrooms based on quartiles, both for mastery goal norms and performance goal norms. As the distribution of mastery goal descriptive norms was negatively skewed in that most classrooms were characterized by quite high mastery goal norms (in line with previous studies, see for instance Ryan & Shim, 2012). Classrooms in the lowest quartile for performance or mastery descriptive norms where indicated as *moderate mastery goal descriptive norm classrooms* and *low performance goal descriptive norm classrooms* (moderate mastery: $M < 4.08$, $N_{\text{classrooms}}$ = 11; low performance: $M < 2.93$, $N_{\text{classrooms}}$ = 11). Classrooms with

descriptive norms in the middle quartiles (25% to 75%) of achievement goals, were referred to as *high mastery goal descriptive norm classrooms* and *moderate performance goal descriptive norm classrooms* (high mastery: $4.08 \leq M \leq 4.41$, $N_{\text{classrooms}} = 24$; moderate performance: $2.93 \leq M \leq 3.47$, $N = 24$). Classrooms in the highest quartile of achievement goals were indicated as *very high mastery goal descriptive norm classrooms* and *high performance goal descriptive norm classrooms* (very high mastery: $M > 4.41$, $N_{\text{classrooms}} = 11$; high performance: $M > 3.47$; $N_{\text{classrooms}} = 11$).

The class-level correlation of achievement goal norms from Wave 1 to Wave 2 was moderate for popularity norms (with $r_{T1\text{-}T2mastery} = .33$, $p = .02$; and $r_{T1\text{-}T2performance} = .40$, $p = .01$), and moderate to high for descriptive norms (with $r_{T1\text{-}T2mastery} = .43$, $p = .01$; and $r_{T1\text{-}T2performance} = .60$, $p = .001$). Correlations between popularity norms and descriptive norms were low and nonsignificant ($r_{mastery} = .05$, $p = 0.77$; $r_{performance} = .23$, $p = .13$). Correlations between mastery norms and performance norms were low for popularity norms and moderate for descriptive norms ($r_{popularity} = .14$, $p = .37$; $r_{descriptive} = .50$, $p < .001$).

Analytic Strategy

Attrition analyses. We performed attrition analyses for students who had partially missing data on the achievement (goal) variables (13.8% in T1 and 12.0% in T2), and we did not find significant or substantial differences between partially missing cases and complete cases on achievement and achievement goals. Little's missing completely at random test produced a normed chi-square (χ^2/df) of 1.48, indicating that the data were likely missing at random and that it was safe to impute missing values on achievement (goal) data (Bollen, 1989). Therefore, to gain statistical power, we estimated missing values for achievement (goal) data in SPSS using the expectation maximization procedure (Gupta & Chen, 2010).

For the friendship nomination data, missing data due to nonresponse were handled through the SIENA missing data method (Huisman & Steglich, 2008), and participants who joined and left the friendship network between time points were treated using the "last observation carry forward" method (Ripley et al., 2016). In this method, for each missing tie variable, the last previous nonmissing value (if any) is imputed; if the previous values are missing as well, the value 0 (referring to no friendship tie) is imputed. Whenever imputed values are used, parameter estimate updates are based on the nonimputed parts of the data. This minimizes the impact of imputations on the results.

RSIENA analyses. Analyses were conducted using longitudinal social network analysis (also called 'stochastic actor-based models'; Snijders, Steglich, & Schweinberger, 2007) with the Simulation Investigation for Empirical Network Analyses (SIENA 4.0–R Version 3.1.2; RSIENA Version 2.8.9) software program. SIENA allows us to examine the extent to which similarity between friends in academic achievement is the result of selection or socialization processes. An assumption of SIENA is that adolescents change their friendship ties and their behaviors in continuous time between the observation moments (i.e., measurement waves) on the basis of individual preferences. At a given moment, students may either change a friendship tie (i.e., create a new tie, drop an existing tie, or maintain a tie) or their behavior (go one step up, one step down, or keep their behavior the same; also called *microsteps*) in response to the current network structure and the behavior of other peers in the network. In this way, SIENA controls for dynamic feedback between behavior change and friendship change, as well as for structural network and individual predictors for changes in friendships and academic achievement. An important assumption of the model is that students have full information about the relationships and behaviors in the network, which is quite realistic in the current study as we examine small class-level networks (in which adolescents spent most of their time at school) and achievement as perceived by peers (not "objective" achievement like GPA). Parameter estimates are derived from iterative simulations using the Robbins-Monro stochastic approximation algorithm (Ripley et al., 2016). For a detailed, more technical explanation of longitudinal social network analyses, we refer to Snijders and colleagues (2007) and Veenstra, Dijkstra, Steglich, and Van Zalk (2013). In the following paragraphs we discuss the parameters we analyzed in our models. See Table S1 in the online supplemental material for further conceptual interpretation of these effects, for information on how the terminology used in this study corresponds to the terminology used in prior RSIENA studies, and for information on how each variable label can be interpreted.

Parameters in the RSIENA model. RSIENA analyses yield parameter estimates related to the network (i.e., structural dynamics and attribute-dependent selection and maintenance dynamics) and behavior dynamics (i.e., influence dynamics and behavioral tendencies). Most of these parameters can be considered as "control parameters," which have to be included to more accurately assess and avoid overestimation of selection and influence dynamics (Snijders, Van de Bunt, & Steglich, 2010). In the following text, we discuss the parameters that are of main interest for testing our hypotheses. See Appendix S3 in the online supplemental material for more details regarding control parameters.

Selection parameters (Figure 1a). To assess the extent to which similarity in achievement among friends is explained by friendship selection processes, we included several selection parameters. The "effect of achievement on friendship nominations received" indicated the extent to which achievement predicted being selected as a friend. Conversely, the "effect of achievement on friendship nominations given" indicates the extent to which achievement predicted the number of friendship nominations given to peers. By including these two parameters, the "similarity-based selection of friends based on achievement" gave a reliable estimate to test our hypotheses about the extent to which adolescents had the tendency to select similarly achieving friends or not, depending on the peer norm.

Next, to assess the direction of friendship selection, we calculated ego-alter maintenance tables (cf. Ripley et al., 2016) that contained the log odds for friendship selection (i.e., formation). These tables indicate whether similarity-based selection takes especially place among higher achieving students or among lower achieving students.

Maintenance parameters (Figure 1a). We examined the extent to which being similar in achievement predicted that a friendship present at one time point would still be present at the next time point (using endowment effects). A positive parameter for similarity-based maintenance of friends indicates that similarity in achievement predicts friendship maintenance, whereas dissimilarity in achievement predicts friendship dissolution (i.e., deselec-

tion). Next, to assess the direction of friendship maintenance, we calculated ego-alter maintenance tables (cf. Ripley et al., 2016) that contained the log odds for friendship maintenance. These tables indicate whether similarity-based maintenance takes place among higher achieving students or among lower achieving students.

Influence parameters (Figure 1b). To assess the extent to which friendship influence on achievement took place, we included the "Friendship influence on achievement" parameter (average similarity). This reflects the tendency of students to change their academic achievement to more closely resemble their friends' average achievement. This tendency could work in the upward or in the downward direction (or remain similar), depending on whether friends display higher or lower levels of achievement than the adolescent does. To assess the direction of friendship influence on achievement, we calculated ego-alter influence tables (cf. Ripley et al., 2016), indicating whether friends influenced adolescents to increase or decrease in achievement over time.

The moderating role of achievement goal peer norms (Figure 1c). We tested whether peer norms at T1 play a role in friendship processes related to academic achievement in four steps. In Step 1, the aforementioned parameters (selection, maintenance, influence and control parameters) were analyzed in RSIENA for all 46 classrooms in multigroup analyses (Ripley et al., 2016).[1] Hence, in this first step (in line with previous studies) the peer norm within the classroom was not taken into account. In Step 2, we performed 12 additional multigroup analyses for all types of classrooms separately (i.e., classrooms with low, moderate, and high performance goal and mastery goal popularity norms; those with low, moderate, high performance goal descriptive norms; and those with moderate, high, and very high mastery goal descriptive norms, respectively). Hence, in total we performed 13 multigroup analyses examining the extent to which friendship processes took place in different class types distinguished by different peer norms. In Step 3, we tested whether there were significant differences between parameter estimates of selection, maintenance, and influence parameters across classrooms with low, moderate, and high norms (and moderate, high, and very high norms for mastery goal descriptive norms) using the following formula: $z = (\beta_a - \beta_b)/\sqrt{(s.e.^2_a + s.e.^2_b)}$, with estimates and β_a and β_b and standard errors $s.e.^2_a$ and $s.e.^2_b$, respectively. This resulted in a z score that under the null-hypothesis of equal parameters has an approximate standard normal distribution (see Steglich, Sinclair, Holliday, & Moore, 2012, p.367; Laninga-Wijnen et al., 2016). We used the significance criterion of $p < .05$.

In Step 4, we assessed convergence of all our models and calculated auxiliary statistics to assess the goodness of fit. Four auxiliary network statistics were assessed: outdegree distribution, indegree distribution, geodesic distance, and triadic census. One auxiliary behavior statistic was assessed: behavioral distribution for achievement. For each auxiliary statistic, the differences between the values in the observed network (summed across the two waves of data) and the simulated values in the model were assessed with the Mahalonobis distance (cf. Ripley et al., 2016) and visually inspected using violin plots.

To facilitate the interpretation of the findings, we calculated odds ratios by taking the exponential function of the parameter estimates (= $exp.[\beta k]$). Odds ratios represent the odds that an outcome will occur given a particular situation, compared with the odds of the outcome occurring in the absence of that situation. For selection and maintenance processes, the odds ratios indicate the odds of adding or retaining someone as a friend relative to the odds for choosing others, conditional on the rest of the model and given the current state of the network. For influence processes, having one additional friend who scores higher (or lower) than oneself increases the odds of an increase (or decrease) in achievement as compared with no change by a factor. For the friendship influence dynamics, we first divided the estimates with the number of answer categories minus one to reflect the effect of a one-unit increase or decrease on the scales. Odds ratios were not calculated for the quadratic shape terms because these are not linear.

Results

Descriptive Statistics

Description of the network and individual variables are presented in Table 1 and Table 2 for classrooms distinguished based on performance goal status norms and mastery goal status norms, respectively. See Appendix S2 in the online supplemental material for a more detailed discussion of these descriptive results. Preliminary analyses indicated that the results were similar for Grade 5 and Grade 6 classrooms. First, we found no significant differences between Grade 5 and Grade 6 classrooms in popularity norms and descriptive norms. Furthermore, the presence and direction of friendship processes related to achievement did not differ significantly across Grade 5 and Grade 6. We also found that the role of peer norms in friendship processes was similar in Grade 5 and Grade 6 classrooms. Therefore, we performed our final analyses on both grades together, in order to gain power. Convergence of all models was good (overall t-ratio for convergence $< .21$), and in one case, one class was omitted from the multigroup analyses in order to get desirable convergence, which did not affect the interpretability of results. The goodness of fit was acceptable or good for all auxiliary statistics in in all classrooms, indicated by a nonsignificant Mahalonis distance and violin plots that indicated that the simulated values did not depart too far from the observed values.

Reported in Table 3 are the controlling variables that do not pertain to our research questions (see also Appendix S3 in the online supplemental material). In the following text, we discuss the main results of interest for testing our hypotheses. As expected,

[1] We used multigroup analyses because our classroom-level networks were rather small which prevented us from obtaining well-converged parameter estimates when analyzing the classrooms separately. Therefore, in line with various previous studies that included rather small classrooms (i.e., Svensson et al., 2012; Delay et al., 2016; Shin & Ryan, 2014a; Logis, Rodkin, Gest, & Ahn, 2013; Weerman, 2011), we combined classrooms and analyzed them simultaneously using multigroup analyses. The multigroup option binds these separate class-level data sets into a large multigroup project, assuming that different data sets are unrelated with one another except for having the same parameter values. In other words, each classroom network is assumed to follow the same rule to evolve, except for the behavioral and network rate functions which are allowed to vary (i.e., class-level variation) within the same multigroup project. In this way, multigroup analyses differ from meta-analyses which take into account class-level variation for each parameter in the model. For more information on multigroup analyses, we refer the reader to the *RSIENA Manual* (see p. 96 and further; Ripley et al., 2016).

Table 1
The Role of Performance Goal Popularity Norms in Changes in Friendship Networks and Achievement

Sample	Low performance goal popularity norms, Mean (*SD*) T1	T2	Moderate performance goal popularity norms, Mean (*SD*) T1	T2	High performance goal popularity norms, Mean (*SD*) T1	T2
Friendship						
Average number of friends	5.37 (2.00)	5.04 (1.24)	5.59 (1.49)	5.33 (1.55)	6.16 (1.90)	6.13 (2.57)
Cohesion in friendship network	.30 (.09)	.28 (.06)	.31 (.09)	.28 (.07)	.31 (.08)	.31 (.09)
Proportion reciprocated friendships	.41 (.08)	.39 (.11)	.46 (.12)	.43 (.10)	.39 (.06)	.41 (.13)
Proportion triadic friendships	.55 (.09)	.57 (.08)	.56 (.09)	.57 (.08)	.56 (.11)	.55 (.09)
Achievement change	T1–T2		T1–T2		T1–T2	
Fraction increased students	18.0%		17.8%		22.2%	
Fraction decreased students	23.1%		17.7%		20.1%	
Fraction stable students	58.9%		64.5%		56.4%	
Friendship change						
Average number of friendship changes	84.36 (33.97)		83.83 (37.27)		76.82 (34.68)	
Proportion of stable friendships	.41 (.08)		.45 (.10)		.41 (.09)	
Friendships emerged	38.45 (17.98)		39.62 (20.81)		54.18 (34.12)	
Friendships dissolved	45.91 (24.68)		44.21 (22.30)		47.55 (27.55)	
Friendships maintained	76.36 (37.92)		81.67 (36.17)		86.36 (39.62)	
N classes	11		24		11	
N students	209		471		221	

Note. Achievement refers to peer-perceived achievement. T1 = Time 1 (fall); T2 = Time 2 (spring).

popularity norms played a role in friendship processes rather than descriptive norms. Therefore, we first display our results on popularity norms.

Popularity Norms and Friendship Selection

Performance goal popularity norms. The similarity-based selection effect was significant in the model with *all classrooms* (OR = 1.35; Table 3, first column; Figure 1a). However, the analyses on classrooms with low, moderate and high performance goal popularity norms separately (Figure 1c), indicated that the parameter for similarity-based selection related to achievement was *significantly positive* in classrooms with high performance goal popularity norms and *significantly negative* in classrooms with low performance goal popularity norms. Moreover, similarity-based selection was significantly more likely in classrooms with high performance goal popularity norms compared with classrooms with moderate popularity norms (z = 2.04, p = .04) and low popularity norms (z = 4.11, p < .001). Also, similarity-based selection was significantly more likely in classrooms with moderate performance goals popularity norms compared with classrooms with low popularity norms (z = 2.97, p = .003). Hence, in high performance goal popularity norm classrooms, students were more likely (by 229% higher odds, which can be seen as a large effect) to select a friend who matched their

Table 2
The Role of Mastery Goal Popularity Norms in Changes in Friendship Networks and Achievement

Sample	Low mastery goal popularity norms, Mean (*SD*) T1	T2	Moderate mastery goal popularity norms, Mean (*SD*) T1	T2	High mastery goal popularity norms, Mean (*SD*) T1	T2
Friendship						
Average number of friends	4.85 (1.33)	4.97 (1.37)	6.07 (1.97)	5.77 (1.95)	5.63 (1.10)	5.24 (1.78)
Cohesion in friendship network	.29 (.06)	.28 (.05)	.31 (.09)	.29 (.08)	.32 (.12)	.30 (.07)
Proportion reciprocated friendships	.47 (.12)	.42 (.10)	.42 (.09)	.41 (.12)	.44 (.11)	.42 (.12)
Proportion triadic friendships	.56 (.09)	.55 (.09)	.56 (.09)	.56 (.08)	.55 (.10)	.56 (.07)
Achievement change	T1–T2		T1–T2		T1–T2	
Fraction increased actors	16.1%		20.1%		19.6%	
Fraction decreased actors	16.3%		20.1%		22.9%	
Fraction stable actors	67.6%		59.8%		57.5%	
Friendship change						
Average number of friendship changes	72.55 (29.00)		97.88 (44.86)		82.91 (36.31)	
Proportion of stable friendships	.45 (.09)		.43 (.10)		.41 (.09)	
Friendships emerged	36.18 (20.42)		46.17 (24.21)		42.18 (28.81)	
Friendships dissolved	36.36 (14.73)		51.71 (29.04)		40.73 (12.42)	
Friendships maintained	61.72 (27.78)		90.25 (38.06)		82.27 (36.79)	
N classes	11		24		11	
N students	207		488		206	

Note. Achievement refers to peer-perceived achievement. T1 = Time 1 (fall); T2 = Time 2 (spring).

Table 3

Performance Goal Popularity Norms and Friendship Dynamics Related to Achievement: RSIENA Multigroup Analyses in All Classes and Classes With Low, Moderate, and High Associations Between Popularity and Performance Goals

	All classes (n = 46)			Low performance goal popularity norms (n = 11)			Moderate performance goal popularity norms (n = 24)			High performance goal popularity norms (n = 11)		
SIENA parameters	*B*	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*
Network dynamics												
Tendency to make friends	−1.69***	.04	.18	−1.71***	.07	.18	−1.77***	.05	.17	−1.60***	.07	.20
Reciprocated friendships	1.03***	.04	2.80	.95***	.09	2.59	1.20***	.06	3.32	.77***	.08	2.16
Transitive group formation	.20***	.01	1.22	.20***	.01	1.22	.21***	.01	1.23	.18***	.01	1.20
Cyclical group formation	−.23***	.01	.79	−.25***	.03	.78	−.24***	.02	.79	−.19***	.02	.83
Selection dynamics												
Same gender (1 = boy) selection	.51***	.03	1.67	.57***	.06	1.75	.46***	.04	1.58	.58***	.05	1.79
Same race selection	.24***	.03	1.27	.34***	.06	1.40	.21***	.04	1.23	.22***	.05	1.25
Effect of achievement on friendship nominations received	.08***	.01	1.08	.11***	$.03^{a}$	1.12	.09**	$.02^{a}$	1.09	.08*	$.03^{a}$	1.08
Effect of achievement on friendship nominations given	.04***	.01	1.04	.13***	$.03^{a}$	1.14	−.01	$.02^{b}$	.99	.06	$.03^{ab}$	1.06
Similarity-based selection of friends	.30*	.16	1.35	−.91*	$.37^{a}$	.40	.36	$.21^{b}$	1.43	1.19***	$.35^{c}$	3.29
Maintenance dynamics												
Similarity-based maintenance of friends	.31*	.15	1.36	1.07*	$.39^{a}$	2.92	.26	$.21^{b}$	1.30	−.15	$.35^{c}$	.86
Influence dynamics												
Achievement linear shape	−.13*	.06	.88	−.32	.15	.73	−.10	.09	.90	−.05	.10	.95
Achievement quadratic shape	.36***	.07		.47*	.19		.36***	.09		.29**	.12	
Friendship influence on achievement	2.49***	.78	2.29	3.51	2.05^{a}	3.22	2.25*	1.09^{a}	2.12	2.26	1.41^{a}	2.12

Note. All models represent separate multigroup analyses. *B* = the unstandardized multinominal logit coefficient. Different superscripts of standard errors (*SE*s) indicate that class types differ significantly from each other in estimate (as computed with z tests). Low, moderate, and high performance goal popularity norms refer to low, moderate, and high class-level associations between popularity and performance goals. Achievement refers to peer-perceived achievement.

* $p < .05$. ** $p < .01$. *** $p < .001$.

own achievement than to select someone with a different achievement. These results are in line with the hypothesis that adolescents in classrooms with higher performance goal popularity norms have an increased tendency to select peers as friends based on similar levels of achievement, and not in line with the alternate hypothesis that adolescents would have an increased tendency to select lower achieving peers as friends in classrooms with salient performance goals.

Next, we calculated ego-alter selection tables to inspect the direction of selection processes in classrooms with high and low performance goal popularity norms (ego-alter tables for moderate performance goal classrooms are available from the corresponding author). In classrooms with high performance goal popularity norms (see Table 4), similarity-based selection especially took place among equally high-achieving peers. Moreover, in low performance goal popularity norm classrooms, ego-alter tables indicate that adolescents had a higher tendency to select lower achieving peers as friends (see Table 4). These findings are generally in line with our hypotheses.

Mastery goal popularity norms. For mastery goal popularity norms (see Table 5), the similarity-based selection effects did not differ significantly from each other (low vs. high mastery popularity norms: z = 0.49, p = .62; low vs. moderate mastery popularity norms: z = 0.26, p = .79; moderate vs. high popularity norms: z = 0.29, p = .77). These results are in line with the hypothesis that mastery goal popularity norms are not strong enough to break down the adolescents' tendency to select peers as friends based on similar levels of achievement, and not with the alternate hypothesis that mastery goals would strengthen friendship selection based on similarity in high achievement. We did not calculate ego-alter tables as none of the selection parameters were significant (data available from the corresponding author).

Popularity Norms and Friendship Maintenance

Performance goal popularity norms. In the model with all classrooms, the friendship maintenance parameter was significant (*OR* = 1.36; see Table 3 and Figure 1a). However, the analyses on classrooms with low, moderate, and high performance goal popularity norms separately (see Figure 1c), indicated that the similarity-based maintenance for achievement was only significantly positive in classrooms with low performance goal popularity norms. Furthermore, maintenance processes based on similarity in achievement took significantly less place in classrooms with high performance goals than in classrooms with low performance goal popularity norms (z = 2.33, p = .02), whereas differences between other types of classrooms were nonsignificant (low vs. moderate performance popularity norms: z = 1.00, p = 32; moderate vs. high performance popularity norms: z = 1.84, p = .07). Hence, in low performance goal popularity norm classrooms, students were more likely (by 192% higher odds, which is a large effect) to maintain a friend who matched their own achievement than to maintain a friend with a different achievement. We calculated ego-alter maintenance tables for low and high performance popularity norms which indicated that in high performance goal popularity norm classrooms, adolescents maintained friendships with peers who were dissimilar in achievement in classrooms; for instance, adolescents with higher achievement had a tendency to maintain lower achieving peers as friends (see Table 4). At the same time, in low performance goal popularity norm classrooms,

Table 4
Likelihood of Peer Selection and Maintenance Based on Achievement in Classes With Low and High Performance Goal Popularity Norms

Individual	Peer 1	Peer 2	Peer 3	Peer 4
	Selection in classrooms with low performance goal popularity norms			
1	−.74	−.33	.09	.50
2	−.30	−.49	−.08	.33
3	.13	−.06	−.25	.16
4	.57	.38	.19	−.01
	Selection in classrooms with high performance goal popularity norms			
1	.52	.21	−.12	−.44
2	.18	.66	.34	.02
3	−.16	.32	.79	.47
4	−.50	−.02	.45	.93
	Maintenance in classrooms with low performance goal popularity norms			
1	.11	−.14	−.39	−.64
2	−.12	.35	.10	−.14
3	−.34	.13	.60	.35
4	−.57	−.10	.37	.84
	Maintenance in classrooms with high performance goal popularity norms			
1	−.29	−.17	−.04	.09
2	−.19	−.16	−.03	.09
3	−.08	−.06	−.03	.10
4	.02	.05	.08	.10

Note. Numbers (1 through 4) in the table reflect the strength of attraction for students to select or to remain friends with certain peers on the basis of their levels of achievement (columns dependent on rows). The values in the cells can be transformed to odds by taking the exponential function (exp.[βk]).

high-achieving peers (rather than low-achieving peers) maintained each other as friends based on similarity in achievement (see Table 4). These findings are in line with the hypothesis that in high performance goal popularity norm classrooms, adolescents remain friends with peers who were dissimilar in achievement and not with the alternate hypothesis that they would remain friends with similarly achieving peers.

Mastery goal popularity norms. Next, the analyses on mastery goal popularity norms indicated that there were no significant differences in maintenance processes between the three types of classrooms (low vs. high mastery popularity norms: $z = 0.33$, $p = .74$; low vs. moderate mastery popularity norms: $z = 0.42$, $p = .67$; moderate vs. high mastery popularity norms: $z = 0.01$, $p = .99$). These findings are in line with the hypothesis that mastery goals are not strong enough to break down the tendency of maintaining similar friends, and not with the alternate hypothesis that mastery goals strengthen adolescents' tendency of selecting higher achieving peers as friends. We did not calculate ego-alter tables as none of the maintenance parameters were significant (data are available from the corresponding author).

Popularity Norms and Friendship Influence

Performance goal popularity norms. In the model with all classrooms, the friendship influence parameter was significant ($OR = 2.29$; see Table 3 and Figure 1b), indicating that, in general, adolescents had a tendency to become similar in academic achievement to their friends. Furthermore, the influence parameter estimates did not differ significantly across classrooms with low, moderate and high associations between popularity and performance goals (Figure 1c), implying that, in contrast to our hypothesis, performance goal popularity norms did not play a significant role in friendship influence on achievement (low vs. high performance popularity norms: $z = 0.51$, $p = .61$; low vs. moderate performance popularity norms: $z = 0.55$, $p = .58$; moderate vs. high performance popularity norms: $z = 0.00$, $p = .996$). We did not calculate ego-alter tables to further inspect the direction of friendship influence as the strength of influence effects did not differ significantly between classes with different types of norms (data are available from the corresponding author).

Mastery goal popularity norms. The analyses separated across classrooms with low, moderate, and high mastery goal popularity norms indicated that the friendship influence parameter was negative and nonsignificant in classrooms with low mastery goal popularity norms. Friendship influence processes occurred in classrooms with moderate mastery goal popularity norms, and particularly in classrooms with high mastery popularity norms, indicating an increase in strength of friendship influence processes as the within-classroom association of popularity with mastery goals increased. The estimate for influence processes did not differ significantly between classrooms with high and moderate mastery goal popularity norms ($z = 1.36$, $p = .18$); but significantly between classrooms with moderate and low mastery goal popularity norms ($z = 2.17$, $p = .03$) and low and high mastery goal popularity norms ($z = 2.42$, $p = .02$). Hence, having one additional friend who scored higher (or lower) than oneself made it more likely to increase (or decrease) in achievement as compared with no change by a factor by 907% higher odds in classrooms with high mastery goal popularity norms, which can be interpreted as a very large effect. This implies that, in line with our hypothesis, the tendency to become similar to friends in achievement increases when the within-classroom association between popularity and mastery goals increases.

We calculated ego-alter tables to further inspect the direction of friendship influence on achievement in high mastery goal popularity norm classrooms (and not in low mastery goal popularity norm classrooms as the influence effect was nonsignificant, available upon request). In these classrooms, the differences in the top rows were larger than in the bottom rows, indicating that in contrast to our hypothesis, students were more likely to decrease in achievement when they had low-achieving friends than to increase in achievement when they had high-achieving friends (see Table 6).

Descriptive Norms and Friendship Dynamics

As expected, descriptive norms did not play a role in the extent to which friendship processes took place within classrooms (see Tables S2 and S3 in the online supplemental material; z scores are available from the corresponding author). Hence, the average aggregated mastery and performance goals within the classroom did not play a role in friendship selection, maintenance, and influence processes with regard to achievement over time. Ego-alter tables are also available from the corresponding author.

Table 5

Mastery Goal Popularity Norms and Friendship Dynamics Related to Achievement: RSIENA Multigroup Analyses in Classes With Low, Moderate, and High Associations Between Popularity and Mastery Goals

	Low mastery goal popularity norms (n = 11)			Moderate mastery goal popularity norms (n = 24)			High mastery goal popularity norms (n = 11)		
SIENA parameters	*B*	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*
Network dynamics									
Tendency to make friends	−1.83***	.08	.16	−1.65***	.05	.19	−1.92***	.08	.15
Reciprocated friendships	1.19***	.11	3.29	1.00***	.06	2.72	1.03***	.10	2.80
Transitive group formation	.27***	.02	1.31	.19***	.01	1.21	.27***	.02	1.31
Cyclical group formation	−.31***	.03	.73	−.21***	.02	.81	−.28***	.03	.76
Selection dynamics									
Same gender (1 = boy) selection	.54***	.06	1.72	.42***	.04	1.52	.64***	.06	1.90
Same race selection	.13*	.06	1.14	.25***	.04	1.28	.33***	.06	1.39
Effect of achievement on friendship nominations received	.06	.03[a]	1.06	.10***	.02[a]	1.11	.08*	.03[a]	1.08
Effect of achievement on friendship nominations given	.02	.03[a]	1.02	.06*	.02[a]	1.06	.03	.03[a]	1.04
Similarity-based selection of friends	.40	.26[a]	1.49	.31	.23[a]	1.36	.18	.36[a]	1.20
Maintenance dynamics									
Similarity-based maintenance of friends	.19	.28[a]	1.21	.34	.23[a]	1.40	.34	.35[a]	1.40
Influence dynamics									
Achievement linear shape	.01	.14	1.01	−.12*	.08	.89	−.28	.19	.76
Achievement quadratic shape	.07***	.16		.34***	.09		.65**	.25	
Friendship influence on achievement	−2.00	1.69[a]	.51	2.26*	.99[b]	2.12	6.93*	3.28[b]	10.07

Note. All models represent separate multigroup analyses. B = the unstandardized multinominal logit coefficient. Different superscripts of standard errors (*SE*s) indicate that class types differ significantly from each other in estimate (as computed with z tests). Low, moderate, and high mastery goal popularity norms refer to low, moderate, and high class-level associations between popularity and mastery goals. Achievement refers to peer-perceived achievement.
$^{*}p < .05$. $^{**}p < .01$ $^{***}p < .001$.

Discussion

The current study investigated the role of achievement goal peer norms in friendship processes related to academic achievement. Our results indicate that the salience of mastery and performance goals within the classroom context, measured in terms of popularity norms, has meaningful implications for the magnitude and direction of these processes. Hence, the extent to which popular peers pursue mastery goals or performance goals has implications for the coevolution between friendships and academic achievement across the school year.

Table 6

Likelihood of Peer Influence on Student's Achievement in Classes with High Mastery Goal Popularity Norms

	Individual achievement			
Peer achievement	1	2	3	4
1	6.04	2.19	−.37	−1.64
2	3.73	4.50	1.94	−.67
3	1.42	2.19	4.25	2.98
4	−.89	−.12	1.94	5.29

Note. Numbers (1 through 4) in the table reflect the strength of friendship influence on certain levels of peer-perceived achievement for the student resulting from the average levels of their friends' achievement (columns dependent on rows). The values in the cells in these tables can be transformed to odds by taking the exponential function (exp.[βk]).

The Moderating Role of Achievement Goal Popularity Norms in Friendship Processes

Selection and maintenance. In line with our expectations, we found that performance goal popularity norms moderated friendship selection and maintenance processes related to achievement. Interestingly, the salience of performance goals had a differential impact on friendship selection and maintenance processes: the higher the performance goal popularity norms, the higher the tendency of adolescents to select similarly achieving peers as friends, and the lower the tendency of adolescents to maintain similarly achieving peers as friends. These results provide valuable insight in the differential impact of performance goal popularity norms on friendship selection and maintenance processes. With regard to selection processes, we found in line with one of our hypotheses that similarity-based selection took place among both low-achieving and (especially) high-achieving students in classrooms with high performance goal popularity norm classrooms. Hence our alternate hypothesis that adolescents would select lower achieving peers as friends (possibly due to self-enhancement perspectives) was not supported. In classrooms with high performance goal popularity norms, it may be useful to select similarly achieving friends for two reasons. First of all, classrooms where performance goals are salient are generally characterized by competition and social comparison, implying that students are highly attuned to interpersonal differences in achievement and academic reputation (Brophy, 2005). It could be theorized that when levels of achievement are similar, comparisons are less threatening for self-worth (Elliot et al., 2011; Festinger, 1954). Second, selecting similarly

high-achieving friends (which took place most often) may be useful in classrooms where performance goals are salient, as adolescents may have an exploitation orientation toward other students, even toward friends (Levy-Tossman et al., 2007; Poortvliet et al., 2007). One can profit from information exchange from similarly high-achieving friends and take advantage of their knowledge and skills (Poortvliet et al., 2007) to reach the goal of outperforming others. Therefore, forming friendships with similarly high-achieving friends may serve the salient goal of achieving superiority over others.

However, with regard to maintenance processes, friendships among similarly high-achieving peers are less likely to last in classrooms where performance goals are made salient by popular peers (compared with classrooms with low performance goal popularity norms). More specifically, in line with one of our alternate hypotheses, adolescents had an increased tendency to maintain friendships with peers who were dissimilar in achievement in classrooms where performance goals were salient. Hence, the alternate hypothesis that adolescents would maintain similarly achieving peers as friends was not supported. This might be due to the fact that, as soon as similarly achieving peers become friends, social comparison may increase because they become closer to each other (and the higher the proximity, the more social comparison may take place; Festinger, 1954). Due to this increased proximity, minor differences in academic functioning may become more visible and threatening (for instance, when one friend receives positive feedback from a teacher whereas another does not, or when one friend scores a slightly higher on a test than the other; see also Sommet et al., 2014; Sommet, Darnon, & Butera, 2015). As a consequence, similarly high-achieving friends may increasingly see each other as a threat toward obtaining the goal of outperforming others, which may result in the dissolution of friendships among these similarly high-achieving peers. As very different others are a less relevant source for comparison, friendships among dis-similar peers may be less threatening at the longer term (Festinger, 1954). Second, it could also be theorized that friendships among similarly high-achieving peers dissolve because the quality of these friendship decreases due to the aforementioned competition or "exploitation practices", which may lead to mutual mistrust, tension, and lower intimacy among friends (Levy-Tossman et al., 2007; Poortvliet et al., 2007).

Next, in line with one of our hypotheses, mastery goal popularity norms did not play a role in similarity-based selection, nor in similarity-based maintenance, related to achievement. Hence, the alternate hypothesis that mastery goals would strengthen friendship selection and maintenance based on similarity in high achievement was not supported. Even though previous studies indicated that social comparison may take place in classrooms where mastery goals are salient (Collins, 1996, 2000), it might be the case that social comparison does not take place based on achievement, but rather based on aspirations and underlying motivation to learn more about a particular topic. Therefore, if social comparison would take place in these classes with salient mastery goals, it might not play a role in friendship selection and maintenance related to achievement. In general, it seems the focus on developing competence and the intrinsic value of learning might not be strong enough to break down the tendency to select and maintain similar friends (similarity-attraction hypothesis, Byrne & Lamberth, 1971). Hence, the attraction to similar peers as friends due to higher levels of perceived trust and predictability (Byrne & Lamberth, 1971) may be important in all classrooms, regardless of the mastery goal popularity norm within the classroom.

Friendship influence. Contrary our hypothesis, we found that performance goal popularity norms did not play a role in the extent to which adolescents have a tendency become similar to their friends in terms of achievement. Even though friendship influence was generally lower when the association between popularity and performance goals was higher, the influence parameter did not significantly diverge across classrooms with different performance goal popularity norms. This finding can be explained as follows: although performance goal popularity norm classrooms may be characterized by less information exchange (i.e., Poortvliet et al., 2009), even among friends (Levy-Tossman et al., 2007), it could be hypothesized that students are highly attuned to any useful or high-quality information within their exchanges with their friends because of their exploitation orientation (Poortvliet et al., 2007). In this way, fewer interactions among students may still have important implications for the extent to which friends may influence each other in achievement over time. Future work that includes examination of the quality and quantity of information exchanged in the classroom could further our understanding of the implications of achievement goal norms for friend processes in the classroom.

Next, in line with our hypothesis, mastery goal popularity norms played a role in the extent to which adolescents became similar to their friends in terms of academic achievement. First of all, the tendency of adolescents to become similar to their friends increased when the association between mastery goals and popularity increased. Adolescents are more susceptible to friendship influence on academic achievement in classrooms where mastery goals are the popularity norm. Prior work indicates that mastery goals yield a cooperative goal structure in which adolescents perceive others as helpers to achieve their goals (Elliot et al., 2016; Karabenick, 2003; Roussel et al., 2011; Ryan & Shim, 2012). It could be hypothesized that this may be associated with useful exchange patterns and elaborated problem solving discussions (Harris, Yuill, & Luckin, 2008) in which adolescents reciprocally share information with each other (Porter, 2005; Poortvliet et al., 2007). Information exchange is the mechanism theorized to underlie peer socialization (Kindermann & Gest, 2009; Ryan, 2000) and our results indicate that when popular students increase the conditions for this mechanism by endorsing mastery goals, socialization is enhanced. Future studies are encouraged to test whether the increased tendency to be influenced by friends in these high mastery goal popularity norm classrooms indeed could be due to higher levels of information exchange.

Second, our results indicate that this increased susceptibility for friendship influence in high mastery goal popularity norm classrooms can be beneficial (in that adolescents' achievement will increase when their friends' achievement is higher on average) or detrimental (in that friends may influence adolescents to become lower in achievement). These unanticipated detrimental effects may be explained in two ways. First, previous studies have found that students with mastery goals are less apt to detect low-quality information when working with others which can hinder task performance (Poortvliet et al., 2007). This may be due students' cooperative mindset (i.e., the inclination to view other students as helpers, even lower achieving students; Porter, 2005). Further, the

salience of mastery goals may enhance a focus on what is interesting, which could distract students from the focus of the task. This finding implies it is important for teachers to provide guidance for productive discussions and help-seeking among students, even when they are focused on mastery goals. Second, our finding could be due to the fact that we measured peer-perceived achievement (i.e., academic reputation) instead of teacher-assigned grades. It could be hypothesized that mastery goal popularity norm classrooms are characterized by higher levels of information exchange among students that provide more opportunities for students to learn more about the academic skills of their classmates compared with in classrooms with less information exchange (i.e., high performance goal popularity norm classrooms). As the school year unfolds, there are more opportunities to see classmates struggle with challenging tasks, which may affect their perception on how well their peers are doing at school. Therefore, especially in these high mastery goal popularity norm classrooms, students may be more highly aware of the struggles and difficulties their fellow-students experience, which may result in a decline in peer-perceived achievement of classmates and friends. Future studies could compare friendship processes related to peer-perceived achievement and teacher-assigned grades in mastery goal popularity norm classrooms to investigate whether potential differences may be due to increased knowledge about each other's difficulties in completing tasks.

Achievement Goal Descriptive Norms and Friendship Processes

As expected, descriptive norms did not play a role in the extent to which friendship selection, maintenance and influence processes take place. First of all, this finding could be due to the fact that descriptive norms are a quite subtle aspect of the environment as they represent average aggregated goals. This does not say much about the valence of a particular behavior, as it might be the case that there is a lot of variation within classrooms regarding these goals, and this variation is not taken into account. Moreover, according to social impact theory, the strength of social forces (in this case, peer norms) is a function of the status of peers, closeness of peers, and number of peers present (Latané, 1981). Descriptive norms only represent the last, quite subtle aspect of this function and hence may not be strong enough to determine social impact (Laninga-Wijnen et al., 2016). Second, descriptive norms represent only the behavioral characteristics of a group, whereas popularity norms refer to corresponding rewards of a group given compliance with the norm (i.e., gaining popularity). Third, descriptive norms place equal weight on all students within the classroom, but not all students may be equally influential. As shown in former studies and in the current study, popular students may be especially influential, as popularity is often more highly desired and more actively pursued by adolescents than by children (LaFontana & Cillessen, 2010), and behaving like popular peers may be an important tool to gain popularity in the peer group (Dijkstra, Cillessen, Lindenberg, & Veenstra, 2010). Descriptive norms also include the behaviors of less popular peers and students may have the tendency to behave opposite to the behaviors of these nonpopular students (see, e.g., Teunissen et al., 2012). Therefore, descriptive norms may be less important for friendship processes. Fourth, in the current study, there was not a high variation in the averages of descriptive norms (especially for mastery goals, which is a common finding in other studies, e.g., Kaplan, Middleton, Urdan, & Midgley, 2002; Patrick, Kaplan, & Ryan, 2011; Régner et al., 2007; Urdan, Midgley, & Anderman, 1998). However, previous studies indicated that even variation at the higher end of the mastery goal scales seemed to matter for academic adjustment and interpersonal relations (i.e., Kaplan et al., 2002; Patrick et al., 2011; Régner et al., 2007; Urdan et al., 1998). Therefore, even though there was not a high variation in average descriptive norms, this variation could still have been predictive of friendship processes.

Limitations and Strengths

Several limitations of the present study need to be acknowledged. First, our reasonably complex model could initially not be identified (convergence problems) in our small classroom-level networks of just 11 to 30 students. Therefore, we combined classrooms with similar levels of peer norms (low, moderate and high) and analyzed them simultaneously using multigroup analyses, which is an approach that is in line with various previous studies that included rather small classrooms (i.e., Delay et al., 2016; Logis et al., 2013; Shin & Ryan, 2014a; Svensson, Burk, Stattin, & Kerr, 2012; Weerman, 2011). Although the use of multigroup analyses increases power and allows for model identification, class-level variation is only considered for some and not all parameters.[1] In this way, multigroup analyses differ from meta-analyses which take into account class-level variation for each parameter in the model. Future studies with larger sample sizes may attempt to replicate our study with meta-analyses, so that class-level variation can be taken into account for all parameters in the model. Moreover, these future studies also could include class-level variables like gender-ratio and educational level, as these variables may play a role in the extent to which achievement goal peer norms are associated with friendship processes on achievement (Anderman & Midgley, 1997; Gherasim, Butnaru, & Mairean, 2013; Shin & Ryan, 2014a).

A second limitation is that we analyzed math and science classrooms without attention to potential differences between subjects (and did not have an adequate design or number of classrooms to do so). Some recent work has conceptualized science and math classrooms as having many similar features that affect adolescents' motivation and engagement similarly (see Fredricks et al., 2016; Wang et al., 2016). However, there are also differences in classroom activities (e.g., doing experiments in science but not math). Future work that assesses peer dynamics, engagement and achievement in both domains for the entire sample could address potential differences.

Third, we did not specifically address potential differences in achievement goals and friendship processes between fifth and sixth grade students due to power limitations. Preliminary analyses revealed no significant differences between Grade 5 and Grade 6 with regard to our research questions (see the Results section). Hence, our findings might imply that the extent to which popular students make achievement goals salient for friendship processes related to achievement might be independent of how often students are together. Indeed, previous studies and theory argue that group dynamics emerge when teacher and students come together each day in classroom, be it for alone hour or for most of the day

(Veenstra & Dijkstra, 2011). Moreover, both grades consist of early adolescents, which experience quite similar levels of hormonal changes and a similar peak in the desire for popularity (Steinberg, 2007). Also, previous studies indicated that peer-perceived achievement operated in middle school math and science classrooms in similar ways as in elementary school classrooms (North & Ryan, 2017). Future studies with larger sample sizes are encouraged to further examine whether achievement goals of the norms of popular peers play a similar role in the coevolution of friendships and achievement in both Grade 5 and Grade 6.

These limitations notwithstanding, this research has several strong and innovative points. First, our study responds to the "context gap" in the current literature on friendship processes regarding academic achievement. Until now, studies investigated achievement-related friendship processes without considering the broader social peer context in which these processes take place. Our study, aimed at capturing the broader social context in terms of achievement goal peer norms (specifically in terms of popularity norms), found that the direction and magnitude of friendship processes is dependent upon the broader social peer context in which they take place. An avenue for future research on the role of peer norms and friendship processes related to achievement would be to analyze whether peer norms play a role in the relative contribution of selection and socialization processes (see, e.g., Rambaran & colleagues, 2016). On the basis of the results of the current study, it could be expected that in classrooms with higher performance goals, selection processes would contribute more to similarity than socialization, whereas in classrooms with higher mastery goals, this would be the opposite.

Second, next to selection processes, we analyzed friendship maintenance processes, which have been rarely studied so far with regard to achievement. Our results indicate the importance of making a distinction between these two processes, as the context (in terms of performance goal popularity norms) may play a differential role in the direction and magnitude of these processes. Hence, future studies are encouraged to make a distinction between maintenance and selection processes related to achievement. Also, we encourage future researchers to take into account the quality (e.g., a 'close' friend or an acquaintance; see for instance Berndt, 1999) of the friendship relationship in examining the role of norms on friendship processes.

Third, an innovative point is that we examined friendship dynamics related to peer-perceived achievement and not to actual grades. The use of peer-perceived achievement as an outcome variable has both practical and theoretical value for the current study. First of all, an important assumption of SIENA is that students have full information about behavior in the network. The use of peer-perceived achievement assures us that we really measure the perception, and thus, the actual information students have on others' behaviors in the network. Second, previous studies indicated that adolescents may especially be influenced (in their friendship choices and in their behavior) by what they think their peers are doing (Bandura, 1986; Helms, Choukas-Bradley, Widman, Giletta, Cohen, & Prinstein, 2014). They may not always be aware of the GPA of other peers, but their close proximity and interactions with classmates may certainly contribute to their perceptions on how well someone is doing at school (Gest et al., 2008). Hence, capturing the perceptions of peers may provide novel intriguing information on how selection, maintenance and influence processes related to achievement takes place.

Contributions and Future Directions

Contributions of our study are twofold. First of all, our research adds to the current field by adopting a social psychology perspective on the role of achievement goals (Doise, 1986; Darnon et al., 2012) and by adequately examining processes of achievement-based friendship selection, maintenance, and socialization with stochastic actor-based modeling. In this way, the current study adds a new dimension to a more social understanding of achievement goals and contributes to our understanding of the interpersonal effects of achievement goals (Darnon et al., 2012). Future studies are encouraged to expand upon the current study to examine whether other types of academic peer norms relate to friendship processes on achievement as well, as there may be a variety of peer norms regarding academic behaviors and attitudes.

Second, the current study examined descriptive norms and popularity norms, and showed that (in line with an increasing number of studies on social adjustment; i.e., Laninga-Wijnen et al., 2016; Rambaran et al., 2013) popularity norms create an important context for the coevolution of friendships and behavior (i.e., achievement in the current study). Our results show that in classrooms where performance goals are endorsed by popular students, this may be detrimental for friendships among peers with similar levels of achievement. Moreover, influence processes are marginal, indicating that there are less possibilities to really learn from each other and to improve skills (everybody on their own island). Classrooms where popular students endorse mastery goals seem to provide an environment in which every student can be successful, but also an environment with certain hazards. Students may profit from interactions with friends who are high-achievers, resulting in similarity in achievement over time. However, we also found that higher achieving students may be disadvantaged by interactions with lower achieving friends (possibly because these friends do not share high-quality information in exchanges). Hence, teachers need to provide guidance and support for students' task-related interactions so that the exchanged information remains of high quality (Poortvliet et al., 2007). The higher susceptibility for peer influence in classrooms where popularity norms make mastery goals salient has potential benefits and drawbacks. Therefore, more studies are needed on the protective factors that could play a role in the direction of friendship influence on achievement in these classrooms with mastery goal popularity norms.

By indicating the importance of popularity norms for friendship processes related to achievement, the current study presents a clarion call to perform more studies in the educational field on the role of popularity norms for academic behaviors and social relations (McCormick & Cappella, 2014). Especially during adolescence, when popularity is such a highly valued characteristic and goal, the norms of popular adolescents may have a profound impact on which academic behaviors are positively valued and reputationally salient within a particular setting (Hartup, 1996). Another interesting area of future research would be to also examine the potential moderating role of individual-level popularity in friendship processes related to achievement in this age-group. This may provide a fuller account of the role of popularity in friendship processes related to achievement in early adolescence. Importantly, in the current study we made a first

step in investigating the role of achievement goal salience in friendship processes, by focusing on the role of performance and mastery goal status norms separately. Our study provides a basis regarding these key relations for future studies that may examine the effect of these popularity norms more in depth. For instance, as performance goals and mastery goals can also form constellations within classrooms (Tuominen-Soini, Salmela-Aro, & Niemivirta, 2011), according to the multiple goals perspective (Pintrich, 2000), and it might be interesting to examine what friendship processes look like in classrooms where both performance and mastery goals are salient, compared with classrooms where either mastery or performance goals are salient or in classrooms where neither achievement goals are salient.

Conclusion

In conclusion, by considering the achievement goals of popular students in classrooms in relation to friendship dynamics across the school year, the present research contributed to the literature on achievement goals as well as friendship processes related to achievement. Classrooms are social places where students are developing friendships and learning, and our results shine light on the complex interplay between social and academic adjustment during early adolescence. For decades, theory and research has given much attention to how teachers affect students' achievement goals and learning outcomes (Ames, 1992; Brophy, 2005). There has been growing recognition in recent years about the role that teachers play in peer dynamics in classrooms (Farmer, Lines, & Hamm, 2011; Gest, Madill, Zadzora, Miller, & Rodkin, 2014). An implication of our findings is that attention to popularity dynamics by teachers is warranted and likely to play a key role in the motivational climate in classrooms of early adolescent students. Teachers receive little to no training in how to manage peer relationships in the classroom. When asked about their efficacy for managing peer relations, both elementary and middle school teachers reported feeling less efficacious about this aspect of their work compared with instruction, motivation and classroom management (Ryan, Kuusinen, & Bedoya-Skoog, 2015). Thus, research and theory to guide professional development supporting teachers in managing peer relationships is an important direction for educational psychology that could advance our understanding of how educators can best support early adolescents' social and academic adjustment.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84,* 261–271. http://dx.doi.org/10.1037/0022-0663.84.3.261

Anderman, E. M., & Midgley, C. (1997). Changes in achievement goal orientations, perceived academic competence, and grades across the transition to middle-level schools. *Contemporary Educational Psychology, 22,* 269–298. http://dx.doi.org/10.1006/ceps.1996.0926

Anderman, E. M., & Wolters, C. (2006). Goals, values, and affect: Influences on student motivation. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 369–348). Mahwah, NJ: Lawrence Erlbaum.

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Clinical and Social Psychology, 4,* 359–373.

Berndt, T. J. (1999). Friends' influence on students' adjustment to school. *Educational Psychologist, 34,* 15–28. http://dx.doi.org/10.1207/s15326985ep3401_2

Bollen, K. (1989). *Structural equations with latent variables.* New York, NY: Wiley. http://dx.doi.org/10.1002/9781118619179

Brophy, J. (2005). Goal theorists should move on from performance goals. *Educational Psychologist, 40,* 167–176. http://dx.doi.org/10.1207/s15326985ep4003_3

Butler, R. (1995). Motivational and informational functions and consequences of children's attention to peers' work. *Journal of Educational Psychology, 87,* 347–360. http://dx.doi.org/10.1037/0022-0663.87.3.347

Byrne, D., & Lamberth, J. (1971). Cognitive and reinforcement theories as complementary approaches to the study of attraction. In B. I. Murstein (Ed.), *Theories of attraction and love* (pp. 59–84). New York, NY: Springer.

Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology, 1,* 659–663. http://dx.doi.org/10.1037/h0022073

Cairns, R. B., Cairns, B. D., Xie, H., Leung, M. C., & Hearne, S. (1998). Paths across generations: Academic competence and aggressive behaviors in young mothers and their children. *Developmental Psychology, 34,* 1162–1174. http://dx.doi.org/10.1037/0012-1649.34.6.1162

Collins, R. L. (1996). For better or worse: The impact of upward social comparison on self-evaluations. *Psychological Bulletin, 119,* 51–69. http://dx.doi.org/10.1037/0033-2909.119.1.51

Collins, R. L. (2000). Among the better ones. Upward assimilation in social comparison. In J. Suls & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 159–171). New York, NY: Kluwer Academic/Plenum Press. http://dx.doi.org/10.1007/978-1-4615-4237-7_9

Crosnoe, R., & Benner, A. D. (2015). Children at School. In M. Bornstein & T. Leventhal (Eds.), *Handbook of child psychology and developmental science* (Vol. 4, pp. 268–304). New York, NY: Wiley.

Darnon, C., Dompnier, B., & Poortvliet, P. M. (2012). Achievement goals in educational contexts: A social psychology perspective. *Social and Personality Psychology Compass, 6,* 760–771. http://dx.doi.org/10.1111/j.1751-9004.2012.00457.x

DeLay, D., Zhang, L., Hanish, L. D., Miller, C. F., Fabes, R. A., Martin, C. L., . . . Updegraff, K. A. (2016). Peer influence on academic performance: A social network analysis of social–emotional intervention effects. *Prevention Science, 17,* 903–913. http://dx.doi.org/10.1007/s11121-016-0678-8

Deutsch, M. (1949). A theory of cooperation and competition. *Human Relations, 2,* 129–152. http://dx.doi.org/10.1177/001872674900200204

Deutsch, M. (1962). Cooperation and trust: Some theoretical notes. In M. R. Jones (Ed.), *Nebraska symposium on motivation* (pp. 275–319). Lincoln, NE: University of Nebraska Press.

Dijkstra, J. K., Cillessen, A. H. N., & Borch, C. (2013). Popularity and adolescent friendship networks: Selection and influence dynamics. *Developmental Psychology, 49,* 1242–1252. http://dx.doi.org/10.1037/a0030098

Dijkstra, J. K., Cillessen, A. H. N., Lindenberg, S., & Veenstra, R. (2010). Basking in reflected glory and its limits: Why adolescents hang out with popular peers. *Journal of Research on Adolescence, 20,* 942–958. http://dx.doi.org/10.1111/j.1532-7795.2010.00671.x

Dijkstra, J. K., & Gest, S. D. (2015). Peer norm salience for achievement, prosocial behavior and bullying: Implications for adolescent school experiences. *The Journal of Early Adolescence, 35,* 1–18. http://dx.doi.org/10.1177/0272431614524303

Dijkstra, J. K., Lindenberg, S., & Veenstra, R. (2008). Beyond the class norm: Bullying behavior of popular adolescents and its relation to peer acceptance and rejection. *Journal of Abnormal Child Psychology, 36,* 1289–1299. http://dx.doi.org/10.1007/s10802-008-9251-7

Doise, W. (1986). *Levels of explanation in social psychology.* New York, NY: Cambridge University Press.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040–1048. http://dx.doi.org/10.1037/0003-066X.41.10.1040

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence. The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist, 48,* 90–101. http://dx.doi.org/10.1037/0003-066X.48.2.90

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52–72). New York, NY: Guilford Press.

Elliot, A. J., Aldhobaiban, N., Kobeisy, A., Murayama, K., Goclowska, M. A., Lichtenfeld, S., & Khayat, A. (2016). Linking social interdependence preferences to achievement goal adoption. *Learning and Individual Differences, 50,* 291–295. http://dx.doi.org/10.1016/j.lindif.2016.08.020

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology, 80,* 501–519. http://dx.doi.org/10.1037/0022-3514.80.3.501

Elliot, A. J., Murayama, K., & Pekrun, R. (2011). A 3 × 2 achievement goal model. *Journal of Educational Psychology, 103,* 632–648. http://dx.doi.org/10.1037/a0023952

Farmer, T. W., Lines, M. M., & Hamm, J. V. (2011). Revealing the invisible hand: The role of teachers in children's peer experiences. *Journal of Applied Developmental Psychology, 32,* 247–256. http://dx.doi.org/10.1016/j.appdev.2011.04.006

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140. http://dx.doi.org/10.1177/001872675400700202

Flashman, J. (2012). Academic achievement and its impact on friend dynamics. *Sociology of Education, 85,* 61–80. http://dx.doi.org/10.1177/0038040711417014

Franke, M. L., Kazemi, E., & Battey, D. (2007). Mathematics teaching and classroom practice. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 225–256). Charlotte, NC: Information Age.

Fredricks, J. A., Wang, M., Schall, J., Hokfkens, T., Parr, A., & Snug, H. (2016). Using qualitative methods to develop a measure of math and science engagement. *Learning and Instruction, 43,* 5–15. http://dx.doi.org/10.1016/j.learninstruc.2016.01.009

Galván, A., Spatzier, A., & Juvonen, J. (2011). Perceived norms and social values to capture school culture in elementary and middle school. *Journal of Applied Developmental Psychology, 32,* 346–353. http://dx.doi.org/10.1016/j.appdev.2011.08.005

Gest, S. D., Madill, R. A., Zadzora, K. M., Miller, A. M., & Rodkin, P. C. (2014). Teacher management of elementary classroom social dynamics: Associations with changes in student adjustment. *Journal of Emotional and Behavioral Disorders, 22,* 107–118. http://dx.doi.org/10.1177/1063426613512677

Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): The association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology, 44,* 625–636. http://dx.doi.org/10.1037/0012-1649.44.3.625

Gherasim, L. R., Butnaru, S., & Mairean, C. (2013). Classroom environment, achievement goals and math's performance: Gender differences. *Educational Studies, 39,* 1–12. http://dx.doi.org/10.1080/03055698.2012.663480

Greenwood, C. R. (1991). Classwide peer tutoring: Longitudinal effects on the reading, language, and mathematics achievement of at-risk students. *Journal of Reading, Writing, and Learning Disabilities International, 7,* 105–123. http://dx.doi.org/10.1080/0748763910070203

Gupta, M. R., & Chen, Y. (2010). Theory and Use of the EM Algorithm. *Foundations and Trends in Signal Processing, 4,* 223–296. http://dx.doi.org/10.1561/2000000034

Harris, A., Yuill, N., & Luckin, R. (2008). The influence of context-specific and dispositional achievement goals on children's paired collaborative interaction. *British Journal of Educational Psychology, 78,* 355–374. http://dx.doi.org/10.1348/000709907X267067

Hartup, W. W. (1996). The company they keep: Friendships and their developmental significance. *Child Development, 67,* 1–13. http://dx.doi.org/10.2307/1131681

Helms, S. W., Choukas-Bradley, S., Widman, L., Giletta, M., Cohen, G. L., & Prinstein, M. J. (2014). Adolescents misperceive and are influenced by high-status peers' health risk, deviant, and adaptive behavior. *Developmental Psychology, 50,* 2697–2714. http://dx.doi.org/10.1037/a0038178

Henry, D., Guerra, N., Huesmann, R., Tolan, P., VanAcker, R., & Eron, L. (2000). Normative influences on aggression in urban elementary school classrooms. *American Journal of Community Psychology, 28,* 59–81. http://dx.doi.org/10.1023/A:1005142429725

Huisman, M., & Steglich, C. E. G. (2008). Treatment of non-response in longitudinal network data. *Social Networks, 30,* 297–309. http://dx.doi.org/10.1016/j.socnet.2008.04.004

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research.* Edina, MN: Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (2005). New developments in social interdependence theory. *Psychological Monographs, 131,* 285–358.

Kaplan, A., Middleton, M. J., Urdan, T., & Midgley, C. (2002). Achievement goals and goal structures. In C. Midgley (Ed.), *Goals, goal structures, and patterns of adaptive learning* (pp. 21–53). Mahwah, NJ: Lawrence Erlbaum.

Karabenick, S. A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology, 28,* 37–58. http://dx.doi.org/10.1016/S0361-476X(02)00012-7

Kindermann, T., & Gest, S. D. (2009). Assessment of the peer group: Identifying naturally occurring social networks and capturing their effects. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships and groups* (pp. 100–117). New York, NY: Guilford Press.

Kiuru, N., Salmela-Aro, K., Nurmi, J.-E., Zettergren, P., Andersson, H., & Bergman, L. (2012). Best friends in adolescence show similar educational careers in early adulthood. *Journal of Applied Developmental Psychology, 33,* 102–111. http://dx.doi.org/10.1016/j.appdev.2011.12.001

LaFontana, K. M., & Cillessen, A. H. N. (2010). Developmental changes in the priority of perceived social status in childhood and adolescence. *Social Development, 19,* 130–147. http://dx.doi.org/10.1111/j.1467-9507.2008.00522.x

Laninga-Wijnen, L., Harakeh, Z., Steglich, C., Dijkstra, J. K., Veenstra, R., & Vollebergh, W. (2016). The norms of popular peers moderate friendship dynamics of adolescent aggression. *Child Development.* Advance online publication. http://dx.doi.org/10.1111/cdev.12650

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36,* 343–356. http://dx.doi.org/10.1037/0003-066X.36.4.343

Levy-Tossman, I., Kaplan, A., & Assor, A. (2007). Academic goal orientations, multiple goal profiles, and friendship among early adolescents. *Contemporary Educational Psychology, 32,* 231–252. http://dx.doi.org/10.1016/j.cedpsych.2006.06.001

Levy-Tossman, I., Kaplan, A., & Patrick, H. (2004). Early adolescents' achievement goals, social status, and attitudes towards cooperation with peers. *Social Psychology of Education, 7,* 127–159. http://dx.doi.org/10.1023/B:SPOE.0000018547.08294.b6

Linnenbrink-Garcia, L., & Patall, E. A. (2016). Motivation. In E. M. Anderman & L. Corno, Eds. *Handbook of educational psychology* (3rd ed., pp. 91–103). New York, NY: Routledge.

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. *Journal of Personality and Social Psychology, 73,* 91–103. http://dx.doi.org/10.1037/0022-3514.73.1.91

Logis, H., Rodkin, P. C., Gest, S. D., & Ahn, H. J. (2013). Popularity as an organizing factor of preadolescent friendship networks: Beyond proso-

cial and aggressive behavior. *Journal of Research on Adolescence, 23,* 413–423. http://dx.doi.org/10.1111/jora.12033

Martin, A. J., Way, J., Bobis, J., & Anderson, J. (2015). Exploring the ups and downs of mathematics engagement in the middle school years. *The Journal of Early Adolescence, 35,* 199–244. http://dx.doi.org/10.1177/0272431614529365

McCormick, M. P., & Cappella, E. (2014). Conceptualizing academic norms in middle school: A social network perspective. *The Journal of Early Adolescence, 23,* 1–14.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology, 57,* 487–503. http://dx.doi.org/10.1146/annurev.psych.56.091103.070258

Midgley, C. (2002). *Goals, goal structures, and patterns of adaptive learning.* Mahwah, NJ: Lawrence Erlbaum.

Midgley, C., Arunkumar, R., & Urdan, T. C. (1996). 'If I don't do well tomorrow, there's a reason': Predictors of adolescents' use of academic self-handicapping strategies. *Journal of Educational Psychology, 88,* 423–434. http://dx.doi.org/10.1037/0022-0663.88.3.423

North, E. A., & Ryan, A. M. (2017). Peer academic reputation in early adolescence: Associations with changes in academic adjustment in math and science. *The Journal of Early Adolescence.* Advance online publication. http://dx.doi.org/10.1177/0272431617692441

Patrick, H., Mantzicopoulos, P., & Sears, D. (2012). Effective classrooms. In K. Harris, S. Graham, & T. Urdan (Eds.), *Educational psychology handbook* (Vol. 2). Washington, DC: APA Books.

Patrick, H., Kaplan, A., & Ryan, A. M. (2011). Positive classroom motivational environments: Convergence between mastery goal structure and classroom social climate. *Journal of Educational Psychology, 103,* 367–382. http://dx.doi.org/10.1037/a0023311

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38,* 109–119. http://dx.doi.org/10.3102/0013189X09332374

Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25,* 92–104. http://dx.doi.org/10.1006/ceps.1999.1017

Poortvliet, P. M., Anseel, F., Janssen, O., Van Yperen, N. W., & Van de Vliert, E. (2012). Perverse effects of other-referenced performance goals in an information exchange context. *Journal of Business Ethics, 106,* 401–414. http://dx.doi.org/10.1007/s10551-011-1005-8

Poortvliet, P. M., & Darnon, C. (2010). Toward a more social understanding of achievement goals: The interpersonal effects of mastery and performance goals. *Current Directions in Psychological Science, 19,* 324–328. http://dx.doi.org/10.1177/0963721410383246

Poortvliet, P. M., Janssen, O., Van Yperen, N. W., & Van de Vliert, E. (2007). Achievement goals and interpersonal behavior: How mastery and performance goals shape information exchange. *Personality and Social Psychology Bulletin, 33,* 1435–1447. http://dx.doi.org/10.1177/0146167207305536

Poortvliet, P. M., Janssen, O., Van Yperen, N. W., & Van de Vliert, E. (2009). The joint impact of achievement goals and performance feedback on information giving. *Basic and Applied Social Psychology, 31,* 197–209. http://dx.doi.org/10.1080/01973530903058276

Porter, C. O. L. H. (2005). Goal orientation: Effects on backing up behavior, performance, efficacy, and commitment in teams. *Journal of Applied Psychology, 90,* 811–818. http://dx.doi.org/10.1037/0021-9010.90.4.811

Rambaran, J. A., Dijkstra, J. K., & Stark, T. (2013). The power of popularity: Status-based influence processes in adolescent risk attitudes. *Journal of Research on Adolescence, 23,* 574–585. http://dx.doi.org/10.1111/jora.12032

Rambaran, J. A., Hopmeyer, A., Schwartz, D., Steglich, C. E. G., Badaly, D., & Veenstra, R. (2016). Academic achievement, truancy, and the status of peers: Network and behavior dynamics during high school. *Child Development.* Advance online publication. http://dx.doi.org/10.1111/cdev.12611

Régner, I., Escribe, C., & Dupeyrat, C. (2007). Evidence of social comparison in mastery goals in natural academic settings. *Journal of Educational Psychology, 99,* 575–583. http://dx.doi.org/10.1037/0022-0663.99.3.575

Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2016). *Manual for RSIENA.* Oxford, UK: Oxford University Press.

Rodkin, P., & Ryan, A. M. (2012). Child and adolescent peer relations in an educational context. In K. Harris, S. Graham, & T. Urdan (Eds.), *Educational psychology handbook* (pp. 363–389). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/13274-015

Roseth, C. J., Johnson, D. W., & Johnson, R. T. (2008). Promoting early adolescents' achievement and peer relationships: The effects of cooperative, competitive, and individualistic goal structures. *Psychological Bulletin, 134,* 223–246. http://dx.doi.org/10.1037/0033-2909.134.2.223

Roussel, P., Elliot, A. J., & Feltman, R. (2011). The influence of achievement goals and social goals on help-seeking from peers in an academic context. *Learning and Instruction, 21,* 394–402. http://dx.doi.org/10.1016/j.learninstruc.2010.05.003

Ryan, A. M. (2000). The peer group as a context for the socialization of adolescents' motivation, engagement, and achievement in school. *Educational Psychologist, 35,* 101–111. http://dx.doi.org/10.1207/S15326985EP3502_4

Ryan, A. M. (2001). The peer group as a context for the development of young adolescent motivation and achievement. *Child Development, 72,* 1135–1150. http://dx.doi.org/10.1111/1467-8624.00338

Ryan, A. M., Gheen, M. H., & Midgley, C. (1998). Why do some students avoid asking for help? An examination of the interplay among students' academic efficacy, teachers' social-emotional role, and the classroom goal structure. *Journal of Educational Psychology, 90,* 528–535. http://dx.doi.org/10.1037/0022-0663.90.3.528

Ryan, A. M., Kuusinen, C. M., & Bedoya-Skoog, A. (2015). Managing peer relations: A dimension of teacher self-efficacy that varies between elementary and middle schoolteachers and is associated with observed classroom quality. *Contemporary Educational Psychology, 41,* 147–156. http://dx.doi.org/10.1016/j.cedpsych.2015.01.002

Ryan, A. M., & Shim, S. S. (2012). Changes in help seeking from peers during early adolescence: Associations with changes in achievement and perceptions of teachers. *Journal of Educational Psychology, 104,* 1122–1134. http://dx.doi.org/10.1037/a0027696

Sandstrom, M. (2011). The power of popularity: Influence processes in childhood and adolescence. In A. H. N. Cillessen, D. Schwartz, & L. Mayeux (Eds.), *Popularity in the peer system* (pp. 219–244). New York, NY: Guilford Press.

Shaw, M. (1981). *Group dynamics (3rd ed.).* New York, NY: McGraw-Hill.

Shin, H., & Ryan, A. M. (2014a). Early adolescent friendships and academic adjustment: Examining selection and influence processes with longitudinal social network analysis. *Developmental Psychology, 50,* 2462–2472. http://dx.doi.org/10.1037/a0037922

Shin, H., & Ryan, A. M. (2014b). Friendship networks and achievement goals: An examination of selection and influence processes and variations by gender. *Journal of Youth and Adolescence, 43,* 1453–1464. http://dx.doi.org/10.1007/s10964-014-0132-9

Snijders, T. A. B., Van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks, 32,* 44–60. http://dx.doi.org/10.1016/j.socnet.2009.02.004

Snijders, T. A. B., Steglich, C. E. G., & Schweinberger, M. (2007).

Modeling the coevolution of networks and behavior. In K. van Montfort, J. Oud, & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences* (pp. 41–71). Mahwah, NJ: Lawrence Erlbaum.

Sommet, N., Darnon, C., & Butera, F. (2015). To confirm or to conform? Performance goals as a regulator of conflict with more-competent others. *Journal of Educational Psychology, 107,* 580–598. http://dx.doi.org/10.1037/a0037240

Sommet, N., Darnon, C., Mugny, G., Quiamzade, A., Pulfrey, C., Dompnier, B., & Butera, F. (2014). Performance goals in conflictual social interactions: Towards the distinction between two modes of relational conflict regulation. *British Journal of Social Psychology, 53,* 134–153. http://dx.doi.org/10.1111/bjso.12015

Steglich, C. E. G., Sinclair, P., Holliday, J., & Moore, L. (2012). Actor-based analysis of peer influence in A Stop Smoking in Schools Trial (ASSIST). *Social Networks, 34,* 359–369. http://dx.doi.org/10.1016/j.socnet.2010.07.001

Steglich, C. E. G., Snijders, T. A. B., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology, 40,* 329–393. http://dx.doi.org/10.1111/j.1467-9531.2010.01225.x

Steinberg, L. (2007). Risk taking in adolescence: New perspectives from brain and behavioral science. *Current Directions in Psychological Science, 16,* 55–59. http://dx.doi.org/10.1111/j.1467-8721.2007.00475.x

Stodolsky, S. S., & Grossman, P. L. (1995). The impact of subject matter on curricular activity: An analysis of five academic subjects. *American Educational Research Journal, 32,* 227–249. http://dx.doi.org/10.3102/00028312032002227

Sumter, S. R., Bokhorst, C. L., Steinberg, L., & Westenberg, P. M. (2009). The developmental pattern of resistance to peer influence in adolescence: Will the teenager ever be able to resist? *Journal of Adolescence, 32,* 1009–1021. http://dx.doi.org/10.1016/j.adolescence.2008.08.010

Svensson, Y., Burk, W. J., Stattin, H., & Kerr, M. (2012). Peer selection and influence of delinquent behavior of immigrant and nonimmigrant youths: Does context matter? *International Journal of Behavioral Development, 36,* 178–185. http://dx.doi.org/10.1177/0165025411434652

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.

Teunissen, H. A., Spijkerman, R., Prinstein, M. J., Cohen, G. L., Engels, R. C., & Scholte, R. H. J. (2012). Adolescents' conformity to their peers' pro-alcohol and anti-alcohol norms: The power of popularity. *Alcoholism, Clinical and Experimental Research, 36,* 1257–1267. http://dx.doi.org/10.1111/j.1530-0277.2011.01728.x

Tuominen-Soini, H., Salmela-Aro, K., & Niemivirta, M. (2011). Stability and change in achievement goal orientations: A person-centered approach. *Contemporary Educational Psychology, 36,* 82–100. http://dx.doi.org/10.1016/j.cedpsych.2010.08.002

Turner, J. C., & Nolen, S. B. (2015). Introduction: The relevance of the situative perspective in educational psychology. *Educational Psychologist, 50,* 167–172. http://dx.doi.org/10.1080/00461520.2015.1075404

Urdan, T., Midgley, C., & Anderman, E. M. (1998). The role of classroom goal structure in students' use of self-handicapping strategies. *American Educational Research Journal, 35,* 101–122. http://dx.doi.org/10.3102/00028312035001101

Urdan, T., & Schoenfelder, E. (2006). Classroom effects on student motivation: Goal structures, social relationships, and competence beliefs. *Journal of School Psychology, 44,* 331–349. http://dx.doi.org/10.1016/j.jsp.2006.04.003

Veenstra, R., & Dijkstra, J. K. (2011). Transformations in adolescent peer networks. In B. Laursen & W. A. Collins (Eds.), *Relationship pathways: From adolescence to young adulthood* (pp. 135–154). Los Angeles, CA: SAGE.

Veenstra, R., Dijkstra, J. K., Steglich, C., & Van Zalk, M. (2013). Network-behavior dynamics. *Journal of Research on Adolescence, 23,* 399–412. http://dx.doi.org/10.1111/jora.12070

Wang, M. T., Fredricks, J., Hofkens, T., & Schall, J. (2016). The math and science engagement scales: Development, validation, and psychometric properties. *Learning and Instruction, 43,* 16–26. http://dx.doi.org/10.1016/j.learninstruc.2016.01.008

Weerman, F. M. (2011). Delinquent peers in context: a longitudinal network analysis of selection and influence effects. *Criminology, 49,* 253–286. http://dx.doi.org/10.1111/j.1745-9125.2010.00223.x

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2016). Development in achievement motivation and engagement. In R. M. Lerner (Series Ed.), C. Garcia & M. Lamb (Eds.), *Handbook of child psychology. Social and emotional development* (7th ed., pp. 657–700). New York, NY: Wiley.

Wright, J. C., Giammarino, M., & Parad, H. W. (1986). Social status in small groups: Individual-group similarity and the social "misfit". *Journal of Personality and Social Psychology, 50,* 523–536. http://dx.doi.org/10.1037/0022-3514.50.3.523

Received May 17, 2016
Revision received April 19, 2017
Accepted April 21, 2017 ■

Call for Papers
A Focused Collection of Qualitative Studies in the Psychological Sciences: Reasoning and Participation in Formal and Informal Learning Environments

Journal of Educational Psychology
Guest Editors: Tanner LeBaron Wallace and Eric Kuo

Reasoning and participation are two central topics of education research in the psychological sciences. Understanding the mechanisms that govern thought and reasoning has long been a core enterprise of educational psychology and, over time, more modern views on learning have promoted participation as a key feature for research—either as a facilitator of learning, a practice to be learned, or as an operationalization of learning itself.

We are pleased to announce a focused collection highlighting qualitative studies of reasoning and participation in formal and informal learning environments. By inviting studies incorporating qualitative methods, we aim to complement the experimental and longitudinal statistical research on these topics that is typically published in this journal. We encourage submission of papers focused on the following (or closely related) topics:

- Student reasoning and/or participation in novel learning environments or activities
- The relations between student reasoning, motivation, identity, and participation
- Student perceptions and meaning-making during participatory experiences
- Dynamic models of student reasoning that are grounded in data
- Explanatory accounts for how and why participation is successful (or not)
- Identifying new goals or targeted outcomes for reasoning or participation

We especially welcome qualitative studies that demonstrate the possibilities for unique discovery afforded by inductive analysis of rich data sources (e.g., real-time recordings of student reasoning, participation, discourse, and physical action, students' meaning-making anchored to particular interactions experienced). This collection will highlight the benefits of qualitative methods for extending and deepening theoretical and empirical understandings of reasoning and participation in both formal and informal learning environments.

The deadline for manuscript submissions is **March 1, 2018**. We invite authors to contact the Guest Editors of this collection, Tanner LeBaron Wallace (twallace@pitt.edu) and Eric Kuo (erickuo@pitt.edu), for discussion on how to maximize alignment between their submissions and this focused collection, though it is not required. Please follow both APA guidelines as well as specific submission criteria for the journal. When submitting manuscripts, please also indicate your intent to submit to this focused collection in the required cover letter.

All manuscripts must be submitted electronically at http://www.editorialmanager.com/edu. In the submission portal, please select the article type "Special Section: Reasoning & Participation – Qualitative." For more information on the *Journal of Educational Psychology*, please visit http://www.apa.org/pubs/journals/edu/.

Instructions to Authors
Journal of Educational Psychology

www.apa.org/pubs/journals/edu

The main purpose of the *Journal of Educational Psychology* is to publish original, primary psychological research pertaining to education across all ages and educational levels. A secondary purpose of the *Journal* is the occasional publication of exceptionally important theoretical and review articles that are pertinent to educational psychology.

Manuscript preparation. Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (6th ed.). Manuscripts may be copyedited for bias-free language (see pp. 70–77 of the *Publication Manual*). Formatting instructions (all copy must be double-spaced) and instructions on the preparation of tables, figures, references, metrics, and abstracts appear in the *Manual*. For APA's Checklist for Manuscript Submission, see www.apa.org/pubs/journals/edu. **Abstract and keywords.** All manuscripts must include an abstract containing a maximum of 250 words typed on a separate page. After the abstract, please supply up to five keywords or brief phrases. **References.** References should be listed in alphabetical order. Each listed reference should be cited in text, and each text citation should be listed in the References. Basic formats are as follows:

Hughes, G., Desantis, A., & Waszak, F. (2013). Mechanisms of intentional binding and sensory attenuation: The role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychologial Bulletin, 139,* 133–151. http://dx.doi.org/10.1037/a0028566

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* Cambridge, MA: MIT Press.

Gill, M. J., & Sypher, B. D. (2009). Workplace incivility and organizational trust. In P. Lutgen-Sandvik & B. D. Sypher (Eds.), *Destructive organizational communication: Processes, consequences, and constructive ways of organizing* (pp. 53–73). New York, NY: Taylor & Francis.

Adequate description of participants is critical to the science and practice of educational psychology; this allows readers to assess the results, determine generalizability of findings, and make comparisons in replications, extensions, literature reviews, or secondary data analyses. Authors should see guidelines for sample–subject description in the *Manual*. Appropriate indexes of effect size or strength of relationship should be incorporated in the results section of the manuscript (see p. 34 of the *Manual*). Information that allows the reader to assess not only the significance but also the magnitude of the observed effects or relationships clarifies the importance of the findings. **Figures.** Graphics files are welcome if supplied in TIFF or EPS format. APA's policy on publication of color figures is available at http://www.apa.org/pubs/authors/instructions.aspx?item=6.

Publication policies. APA policy prohibits an author from submitting the same manuscript for concurrent consideration by two or more publications. APA policy regarding posting articles on the Internet may be found at www.apa.org/pubs/authors/posting.aspx. In addition, it is a violation of APA Ethical Principles to publish "as original data, data that have been previously published" (Standard 8.13). As this is a primary journal that publishes original material only, APA policy prohibits publication of any manuscript or data that have already been published in whole or substantial part elsewhere. Authors have an obligation to consult journal editors concerning prior publication of any data on which their article depends. In addition, APA Ethical Principles specify that "after research results are published, psychologists do not withhold the data on which their conclusions are based from other competent professionals who seek to verify the substantive claims through reanalysis and who intend to use such data only for that purpose, provided that the confidentiality of the participants can be protected and unless legal rights concerning proprietary data preclude their release" (Standard 8.14). Authors must have available their data throughout the editorial review process and for at least 5 years after the date of publication.

Masked review policy. The *Journal* has a masked review policy, which means that the identities of both authors and reviewers are masked. Every effort should be made by the authors to see that the manuscript itself contains no clues to their identities. Authors should never use first person (*I, my, we, our*) when referring to a study conducted by the author(s) or when doing so reveals the authors' identities, e.g., "in our previous work, Johnson et al., 1998 reported that. . . ." Instead, references to the authors' work should be in third person, e.g., "Johnson et al. (1998) reported that . . .." The authors' institutional affiliations should also be masked in the manuscript. Authors submitting manuscripts are required to include in the cover letter the title of the manuscript along with all authors' names and institutional affiliations. However, the first page of the manuscript should omit the authors' names and affiliations, but should include the title of the manuscript and the date it is submitted. Responsibility for masking the manuscript rests with the authors; manuscripts will be returned to the author if not appropriately masked. If the manuscript is accepted, authors will be asked to make changes in wording so that the paper is no longer masked. Authors are required to state in writing that they have complied with APA ethical standards in the treatment of their sample, or to describe the details of treatment. A copy of the APA Ethical Principles may be obtained at www.apa.org/ethics/ or by writing the APA Ethics Office, 750 First Street, NE, Washington, DC 20002-4242. APA requires authors to reveal any possible conflict of interest in the conduct and reporting of research (e.g., financial interests in a test procedure, funding by pharmaceutical companies for drug research). Authors of accepted manuscripts will be required to transfer copyright to APA.

Permissions. Authors of accepted papers must obtain and provide to the editor on final acceptance all necessary permissions to reproduce in print and electronic form any copyrighted work, including test materials (or portions thereof), photographs, and other graphic images (including those used as stimuli in experiments). On advice of counsel, APA may decline to publish any image whose copyright status is unknown.

Supplemental materials. APA can place supplementary materials online, which will be available via the published article in the PsycARTICLES database. To submit such materials, please see www.apa.org/pubs/authors/supp-material.aspx for details. Authors of accepted papers will be asked to work with the editor and production staff to provide supplementary materials as appropriate.

Submission. Authors should submit their manuscripts electronically via the Manuscript Submission Portal at www.apa.org/pubs/journals/edu/index.aspx (follow the link for submission under Instructions to Authors). General correspondence may be addressed to the editorial office at CJohnson@apa.org.